How people make adaptive decisions with (the help of) others: Studies from an ecological rationality perspective

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

Dr. rer. nat. (doctor rerum naturalium)

im Fach Psychologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von
Dipl.-Psych. Juliane Eva Kämmer

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Gerd Gigerenzer, Max-Planck-Institut für Bildungsforschung
2. Prof. Dr. Elke van der Meer, Humboldt-Universität zu Berlin
3. Prof. Dr. Rocío García-Retamero Imedio, Universidad de Granada

Eingereicht am: 14.12.2012
Disputation am: 30.07.2013

Acknowledgements

This dissertation is the outcome of research I have developed at the Center for Adaptive Behavior and Cognition (ABC), at the Max Planck Institute for Human Development, as a fellow of the International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World (Uncertainty School).

First and foremost, I am most indebted to my supervisor Wolfgang Gaissmaier, who has been a great advisor in all (research) matter, and always been approachable. He was an indispensable knowledge source of academic thinking and relevant literature. It was a privilege to work with him and to learn from him. He had always a good idea and recommendation when I was stuck, and provided me with emotional support.

I am grateful to Gerd Gigerenzer for providing the opportunity for me to work at ABC—an exceptionally inspiring and open-minded research environment. Being at ABC and having had the opportunity to meet numerous extremely interesting researchers there and at great conferences provided me with an exiting (research) time. My experience at ABC has certainly set high standards for future working environments.

The time at ABC would not have been as valuable as it was without my colleagues. I would therefore like to thank all of them for creating a friendly and supportive working atmosphere. They have helped to make this dissertation possible.

Moreover, I am thankful to the Uncertainty School organizers, in particular Susanne Büchner, for taking care of all (little) problems and organizing family-like summer schools and workshops.

I am thankful to my co-authors with whom I enjoyed collaborating very much and of whom I learned a lot. Many thanks to Gregor Caregnato and Jann Wäscher for their help with collecting the data, to Uwe Czienskowski for his help with experimental programming, and to Christian Elsner for his computer support.

I would like to thank Claudia Walther for her friendship and ongoing (and last minute) support and optimistic vibes. I am grateful to my friends and family, especially to my parents, Sabine and Jörg Kämmer, my sister Henriette, and Stephan, my love, for their unconditional support and encouragement.

English Summary

This dissertation is an investigation from an ecological rationality perspective of how people make decisions in a social world. It focuses on two exemplary social contexts: situations in which people collectively make decisions in small groups and in which people use another person's advice. Of particular interest in all the reported studies were the questions of what environmental factors influence the use and performance of different decision strategies, and how they do so. The studies thus were aimed at linking the framework of ecological rationality with research on group decision making and advice taking, respectively, to derive new insights for the related research streams.

A *first project* compared the performances of individuals and two-person groups, or dyads, in a strategy-learning task. The task was to learn with the help of feedback the most adaptive strategy for a given task environment. One environment favored the take-the-best strategy, that is, relying on the best discriminating cue and ignoring the rest; to perform well in the second environment required using the weighted additive strategy, which weights and adds all available cues. Results show that both individuals and dyads learned to select the most appropriate strategy over time, with a steeper learning rate in dyads when take-the-best was adaptive.

A *second project* investigated whether small groups apply decision strategies conditional on the group's composition in terms of task-relevant features. The focus here was on the recognition heuristic, so the task-relevant features that influenced the potential performance of group strategies were the validities of the group members' recognition and knowledge. Results of an experiment with 3-member groups working on a paired-comparison task support the hypothesis that groups indeed adaptively apply the strategy that leads to the highest theoretically achievable performance.

A *third project* investigated the impact of task difficulty on the use of advice from an ecological rationality perspective. Task difficulty was shown to influence statistical properties of the environment, which, in turn, were shown to determine the theoretical accuracy of choosing and averaging, two prominent advice-taking strategies. Results of an experiment also showed that people used different advice-taking strategies depending on the difficulty of the task.

Deutsche Zusammenfassung

Diese Dissertation beschäftigt sich mit der Frage, wie Menschen Entscheidungen in sozialen Kontexten treffen. Es werden Entscheidungen in zwei exemplarischen Kontexten aus der Perspektive der ökologischen Rationalität untersucht: Situationen, in denen Menschen Entscheidungen zusammen mit anderen in kleinen Gruppen treffen, und Situationen, in denen sie Rat suchen. Von besonderem Interesse sind in allen Studien der Dissertation die Fragen, wie und welche Umweltfaktoren die Verwendung und Güte von verschiedenen Entscheidungsstrategien beeinflussen. Die Studien haben zum Ziel, den Forschungsrahmen der ökologischen Rationalität mit der Gruppenforschung und Literatur zum Thema „Ratgeben" (advice taking) zu verknüpfen, um für die jeweiligen Forschungsstränge neue Erkenntnisse zu gewinnen.

In einem ersten Projekt wurden die Leistungen von Einzelpersonen und Zweiergruppen in einer Strategielernaufgabe miteinander verglichen. Die Aufgabe war, mit Hilfe von Feedback, die Strategie zu lernen, die adaptiv in Bezug auf die Struktur der Umwelt war. Dabei gab es in der ersten Aufgabenumwelt eine adaptive Strategie, die auf den besten diskriminierenden Cue setzt und die weniger validen Informationen ignoriert (take-the-best), wohingegen die adaptive Strategie in einer zweiten Aufgabenumwelt darauf basierte, alle vorhandenen Informationen zu verrechnen (weighted additive). Die Ergebnisse zeigen, dass sowohl Einzelpersonen als auch Gruppen die jeweils beste Strategie gelernt haben, wobei Gruppen einen schnelleren Lernerfolg zeigten, wenn take-the-best adaptiv war.

Ein zweites Projekt untersuchte, ob kleine Gruppen Entscheidungsstrategien verwenden, die auf ihre Zusammensetzung hinsichtlich aufgabenrelevanter Faktoren abgestimmt sind. Diese Frage wurde mit einer Paarvergleichsaufgabe untersucht, in der die Verwendung der Rekognitionsheuristik eine große Rolle spielt. Die aufgabenrelevanten Faktoren, die die Güte verschiedener Gruppenstrategien beeinflussten, waren dadurch die Validität der Wiedererkennung und des Wissens der Gruppenmitglieder. Ergebnisse eines Experiments mit 3-Personen-Gruppen unterstützten die Hypothese, dass Gruppen dazu in der Lage sind, den Strategien zu folgen, die am erfolgversprechendsten sind.

Ein drittes Projekt untersuchte den Einfluss von Aufgabenschwierigkeit auf die Verwendung von Ratschlägen aus einer ökologisch rationalen Perspektive. Dazu konnten wir zeigen, dass sich Aufgabenschwierigkeit in verschiedenen statistischen Merkmalen der Umweltstruktur niederschlägt, die wiederum die potentielle Güte von zwei häufig verwendeten Strategien (mitteln und auswählen) beeinflussten. Zudem zeigten die Ergebnisse eines Experiments, dass Personen ihre Strategien, mit denen sie Ratschläge mit ihrer eigenen Meinung integrieren, auf die Aufgabenschwierigkeit abstimmen.

Table of Content

Chapter 1

Introduction

In our daily lives, we usually make decisions within a social context. Together with our friends, we decide where to go for a weekend trip. We seek advice from our parents on how to raise a child. Similarly, in the context of work, we seldom make decisions on our own. Take academia as an example. A quick survey of the more than 600 research reports published in *Science* in 2012 reveals that less than 10% were written by a single author, while more than 90% were coauthored by at least one other researcher. Leaving aside alternative reasons for coauthorship, research is an inherently collaborative endeavor. As one’s own resources and time are limited, information sharing and joint decision making are necessary for planning, conducting, and writing up studies. In addition, academia is a field in which giving and taking advice are common: More senior and experienced researchers instruct and supervise students working on their theses or recommend ways to cope with data analysis problems, for example.

Decision making has been found to take place most often within a group of people (Kerr & Tindale, 2004) or under the advice of another person (Bonaccio & Dalal, 2006). This dissertation explores—within the context of the larger research program on decision making under uncertainty—how people make decisions under uncertainty in groups and with the help of others and, more generally, how the environment at large shapes behavior. More precisely, the studies presented here focus on the following three research questions: (1) How do people make decisions within a social context? (2) Does social decision making differ from individual decision making? (3) What environmental factors influence the use and performance of strategies employed by groups for making decisions or by individuals for integrating advice?

In three distinct projects conceived for publication as stand-alone articles, I explored these questions and aimed to link the concept of ecological rationality with research on group decision making and advice taking, respectively, to derive new insights for each related research stream. Before I successively report on the research projects in Chapters 2 to 4, in this Introduction I briefly summarize the central ideas of ecological rationality, the common starting point of the three studies. This is followed by an outline of the crucial importance of studying decision making within a social context and an overview of the three studies, including the specific focus of each of the three projects.

The Concept of Ecological Rationality and a Key Research Gap

Human rational behavior ... is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor.

Simon, 1990, p. 7

Central to the concept of ecological rationality is the question of what specific decision strategies fit to particular environments, that is, are successful in those environments, and how humans select them (Gigerenzer, Todd, & the ABC Research Group, 1999; Todd, Gigerenzer, & the ABC Research Group, 2012). These are the two blades of Herbert Simon's (1990, p. 7) scissors. The most successful strategies are often *heuristics*, or simple rules of thumb.

Heuristics are understood here as decision strategies that ignore some of the available information. They are fast and frugal and exploit the structure of the environment. A strategy is considered *ecologically rational* to the degree that it matches the environmental structure. The recognition heuristic is a prototypical example of a heuristic (Goldstein & Gigerenzer, 1999; 2002; Gigerenzer & Goldstein, 2011). It predicts that in a binary choice task in which one, but not the other, option is recognized, the recognized option will score higher on a criterion. The heuristic works well for tasks where recognition is highly correlated with the criterion. Another example is take-the-best (Gigerenzer & Goldstein, 1999). Take-the-best sequentially searches for information starting with the most valid cue, checking its values for the options at hand, and stopping the search as soon as a discriminating cue is found. The option with the higher value is then chosen. This heuristic benefits from a noncompensatory cue structure, that is, it performs well when the distribution of cue validities is highly skewed or dispersed and/or when cues are highly redundant (Dieckmann & Rieskamp, 2007; Hogarth & Karelaia, 2006, 2007).

These two examples illustrate the basic assumption of the concept of ecological rationality: that humans possess a repertoire of heuristics suited for solving tasks in different domains. This repertoire has been termed the "adaptive toolbox" (Gigerenzer & Selten, 2001; Gigerenzer et al., 1999). No single strategy is assumed to be universally superior (Gigerenzer & Gaissmaier, 2011).

What heuristics does the adaptive toolbox include? Todd and Gigerenzer (2012) gave an overview of a dozen heuristics (see their Table 1-1, pp. 9–10) assumed to be part of the toolbox. Further, additional social heuristics were listed by Hertwig and Herzog (2009, see their Table 1, pp. 684–685). Though even more heuristics are believed to exist, not all are presumed to be in the toolbox of every person at any one time. Instead, its contents are subject to influences of individual learning and experience as well as environmental changes.

Research of the past two decades has provided ample evidence that humans indeed select heuristics in an adaptive way when facing preferential choice (Payne, Bettman, & Johnson, 1993) and inductive inference (e.g., Bröder & Schiffer, 2003, 2006; Rieskamp & Otto, 2006) tasks. Most of this research, however, has focused on individuals (for an exception see Reimer & Katsikopoulos, 2004). The three studies presented in the following chapters should thus complement existing research on the adaptive use of decision strategies by specifically extending the investigation to social settings, which has been a key perspective missing from the current literature.

The Practical and Theoretical Importance of Studying Social Settings

The study of decision making in social settings is motivated by both its practical relevance as well as its theoretical importance. On the practical side, meetings, collaborations, and teamwork are ubiquitous not only in academia, as in the examples given above, but in almost every organizational context (Meyer, Shemla, & Schermuly, 2011; Salas, Cooke, & Rosen, 2008). Moreover, even when making decisions on their own, people often seek advice from (preferably more experienced) others.

One core reason for the prevalence of teamwork is that teams have the potential to outperform people working alone, because they are able to pool information and combine multiple perspectives in an increasingly specialized and complex world (Larson, Foster-Fishman, & Keys, 1994; Stasser, 1992). Interdisciplinary research, for example, is based on the idea that multiple experts from different disciplines pool their knowledge in order to draw a more comprehensive picture of a research issue than any expert alone could do. More generally, it has been observed that researchers frequently generate hypotheses in groups (e.g., Dunbar, 1977). Another advantage in organizations and in society at large is that group decisions often enjoy greater legitimacy and acceptance than individual decisions (Allen & Hecht, 2004).

Yet, small-group research has revealed a number of potential disadvantages of groups, such as process losses due to coordination difficulties (Steiner, 1972) or distraction (Baron, 1986). Take brainstorming as an example. Originally it was thought that brainstorming explicitly benefits from group settings: Osborn claimed that "the average person can think up twice as many ideas when working in a group than when working alone" (if his four rules of brainstorming were adhered to; Osborn, 1957, p. 229). Decades of research on brainstorming has not, however, supported this claim (for a meta-analysis see Mullen, Johnson, & Salas, 1991). Indeed the conclusion reached today holds that effective brainstorming is achieved by the combination of an individual and a group brainstorming phase (e.g., Stroebe, Nijstad, & Rietzschel, 2011).

The results from this field of research imply that there is no consensus in favor of or against group work. Distinguishing the conditions under which groups perform well from those when it is better to assign tasks to individuals is, obviously, of great practical importance. In fact, it paves the way for intelligent design of social systems, by informing the building of teams and organizations so as to maximize the potentials of both groups and individuals. Taking an ecological rationality perspective promises to contribute to a differentiated understanding of these complex issues.

The second motivating force for studying decision making in social settings is of a theoretical nature. The literature on group decision making in social psychology and on individual decision making in cognitive psychology address very similar questions; yet these two streams of research so far remain largely unconnected. The three studies reported in this dissertation show how fruitful it can be to combine the literature of the two disciplines in order to shed light on specific questions, such as whether a group's selection of appropriate decision strategies depends on its composition in terms of certain task-relevant features (Chapter 2).

Beyond contributing to finding answers to such specific open research questions, revealing the parallels between the two corresponding research streams in social and cognitive psychology can inspire theory building in both fields. Obvious similarities exist, for example, in the heuristics used by individuals, such as take-the-best and tallying, and formal group decision-making strategies, such as the best member rule and the simple majority rule (cf. Reimer & Hoffrage, 2012). Reimer and Hoffrage (2012) suggested that "these formal similarities between the individuals' strategies and social combination rules allow us to extrapolate some of the lessons from the ecological rationality of decision strategies for individuals to the ecological rationality of social combination rules" (p. 356). Comparing the formalization of heuristics in terms of their ecological rationality and basic building blocks—decision rules governing search, stopping search, and deciding—with tests of formal models of decision making may inspire further research on the following two important questions. First, can equivalences be identified in the decision rules used by individuals and groups? Second, what environmental structures, including group composition, make different group decision rules ecologically and socially rational?

In a similar vein, recent advances in research on the use of advice have shown how applying a new framework can provide fresh insight into seemingly established findings: Taking an ecological rationality perspective and using individual-level analyses, Soll and Larrick (2009) revealed that humans use mainly two distinct strategies when integrating advice—choosing and averaging—and do *not* apply one overall strategy, namely, adjusting by 30% toward the advice, as previous research had concluded (Harvey & Fischer, 1997; Yaniv, 2004; Yaniv & Kleinberger, 2000). Soll and Larrick's proposed model analytically derived predictions about the relative performance of these two strategies. In Chapter 4, this model is discussed and applied in detail.

Finally, one area that might benefit from integrating the two literatures of social and cognitive psychology is the study of differences between individuals' and groups' information processing and decision making (Hinsz, Tindale, & Vollrath, 1997). Conceptualizing groups as information-processing entities where cognition is distributed across individuals (De Dreu, Nijstad, & van Knippenberg, 2008; Hinsz et al., 1997; Larson & Christensen, 1993) has enabled researchers to "assess if, when, and how group information processing is similar to, or different from, its individual-level counterpart" (Hinsz et al., 1997, p. 44). It could be informative, for example, to study the similarities and differences between groups and individuals when they are learning a decision strategy that fits to the environment (Chapter 2). This could fill a research gap recognized only recently by Reimer and Hoffrage (2012): "more empirical studies are needed to determine the extent to which groups do (or can) use heuristics when forming a decision. … We also need to explore how sensitive groups and their individual members are to particular structures of their environment" (pp. 358–359).

Thus it is clear that extending research on decision making from the individual to the group level and applying an ecological rationality perspective to the study of decision making in social settings have both practical relevance and the potential to inspire theories in social and cognitive psychology. In the remainder of this section I provide an introduction to the chapters to follow that form the core of this dissertation. The main topics are briefly outlined and the specific facets of ecologically rational behavior they cover are highlighted.

Overview of the Studies in this Dissertation

The first study, reported in Chapter 2, explored the question of how the environment shapes decision strategies used in a multiple cue comparison task. The issue has been studied before in individuals (e.g., Bröder, 2003; Bröder & Schiffer, 2003, 2006; Rieskamp & Otto, 2006) and is extended here to the group level. More specifically, groups and individuals were confronted with an unfamiliar task and it was observed how quickly and how well they adapted their decision strategy over repeated trials. The task environment was manipulated so that it favored either the take-the-best strategy (Goldstein & Gigerenzer, 1999) or the weighted additive strategy (Dawes, 1979) in two experiments—the first a between-subjects design and the second a within-subject design. The weighted additive strategy searches for all cues per option, multiplies each cue value by its weight, and finally selects the option with the larger weighted sum. Weighted additive performs well in environments with a compensatory cue structure, that is, where cues are similarly valid and nonredundant, thus adding new evidence when acquired.

In the second study, reported in Chapter 3, the environmental structure was held constant and the focus was on how the cognitive capabilities of multiple human minds and the groups' specific compositions influenced strategy use and performance. Unlike the study in Chapter 2, which dealt with "inferences from givens," this study explored "inferences from memory" (Gigerenzer & Todd, 1999). In an experiment, study participants met in three-member ad-hoc groups and discussed which of two German companies had a higher market capitalization. The recognition heuristic plays an important role in such tasks when people work individually (e.g., Goldstein & Gigerenzer, 2002; Marewski & Schooler, 2011); here I investigated the adaptive use of recognition by small groups (cf. Reimer & Katsikopoulos, 2004). How would groups value the contribution of members who used the recognition heuristic and thus based their decisions on a lack of knowledge? Would groups select decision strategies that took advantage of the composition of the group in terms of the knowledge/recognition distribution across members?

The final study, in Chapter 4, examined how task difficulty influences the effectiveness and use of different advice-taking strategies. While the concept of ecological rationality has shaped much of the research in the area of judgment and decision making over the last decades, this approach is relatively new in the advice-taking literature. In this third study I focused on the judge–advisor system (JAS) in quantitative estimations. In the JAS, judges give an initial estimate, receive advice in the form of another person's estimate, and then have a chance to revise their first estimate. Imagine that you (the judge) have to estimate the number of days you and your colleague will need to submit a paper together. Your colleague (the advisor) also gives an estimate. When asked for a final estimate, you could either average the two initial estimates, choose your own estimate, or choose your colleague's.

Applying ecological rationality to the study of advice taking, in this study I approached the questions of whether tasks of different levels of difficulty require different strategies and, if so, whether humans act accordingly. First, I explored the statistical properties of easy, intermediate, and difficult task environments and determined their impact on the theoretical performance of different advice-taking strategies, such as averaging and choosing. Thereafter I derived predictions from Soll and Larrick's (2009) probability, accuracy, redundancy (PAR) model. Last, I tested if humans use averaging and choosing to different extents contingent on environmental properties.

In sum and as the core common thread, the three studies of this dissertation explored heuristic decision making in social contexts from an ecological approach. Chapter 5 contains a summary and discussion of the main findings and prospects for further research.

Chapter 2

The environment matters: Comparing individuals and dyads in their adaptive use of decision strategies[1]

Kämmer, J. E., Gaissmaier, W., & Czienskowski, U.

Abstract

Individuals have been shown to adaptively select decision strategies depending on the environmental structure. Two experiments extended this research to the group level. Participants (*N* = 240) worked either individually or in two-person groups, or dyads, on a multi-attribute paired-comparison task. They were randomly assigned to two different environments that favored one of two prototypical decision strategies—weighted additive or take-the-best (between-subjects design in Experiment 1 and within-subject design in Experiment 2). Both individuals and dyads learned to select take-the-best over time when it was adaptive, with a steeper learning rate in dyads. Selecting weighted additive when it was adaptive was equally likely for individuals and dyads. Information search data support this finding. Analyses of nominal dyads indicate that real dyads performed at the level of the best individuals.

[1] A version of this chapter was published in 2013 in *Judgment and Decision Making, 8,* 299-329.

Introduction

Imagine a group of geologists searching for profitable oil-drilling sites for an oil company. Before this group can pick one of several possible sites, it has to decide *how* to make this decision. First, it needs to decide what information to search for and in what order. Different methods are available for inferring the quality of the available sites, such as chemical and seismic analyses, which differ in their success rate. Second, the group needs to decide when to stop searching for information and, third, how to integrate the pieces of information to make a decision. For example, it could commission all available analyses and weight and add the results. Alternatively, it could proceed sequentially, starting with the most successful method and deciding as soon as one result clearly favors one site.

This example illustrates the idea that decision makers can choose from a repertoire of different decision strategies, for which Gigerenzer, Todd, and the ABC Research Group (1999) have coined the term "adaptive toolbox." This idea goes back to Herbert A. Simon (1956) who saw cognition as an adaptation to the environment. Different environments require the use of different decision strategies to be successful, as no single strategy will be universally superior (Gigerenzer & Gaissmaier, 2011; yet there is a lively debate about whether a Bayesian approach to cognition could be such a universal strategy, see e.g., Jones & Love, 2011; for comments see Bowers & Davis, 2012a; Bowers & Davis, 2012b; Griffiths, Chater, Norris, & Pouget, 2012). A strategy is considered *ecologically rational* to the degree that it matches the environmental structure. The important questions are whether people are good at deciding how to decide, and how they do so. This fundamental problem is known in the literature as the strategy selection problem (e.g., Payne, Bettman, & Johnson, 1988, 1993; Rieskamp & Otto, 2006; for alternative accounts see Newell & Lee, 2011; for a debate see Glöckner, Betsch, & Schindler, 2010; Marewski, 2010).

Within the existing literature on adaptive strategy selection in humans (e.g., Bröder, 2003; Christensen-Szalanski, 1978, 1980; Marewski & Schooler, 2011; Payne et al., 1988, 1993; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006), most of the research has focused on adaptive decision making in individuals (for rare examples see Kämmer, Gaissmaier, Reimer, & Schermuly, 2012; Reimer & Katsikopoulos, 2004).

Many decisions in real life, however, are made in a social context, for example, under the advice of another person (e.g., Bonaccio & Dalal, 2006) or in a group of people (Kerr & Tindale, 2004; Levine & Smith, 2012). In fact, teams are ubiquitous in all sectors of organizations today, as, for example, in the healthcare system or aviation (Manser, 2008; Waller, 1999). Reasons for this prevalence are mainly seen in (a) their potential superiority to individuals as they can combine multiple perspectives, areas of expertise, and resources to work on complex problems (Larson, Foster-Fishman, & Keys, 1994; Stasser, 1992) and (b) their large potential for adaptation to a dynamic environment (Burke, Stagl, Salas, Pierce, & Kendall, 2006; Randall, Resick, & DeChurch, 2011). Because of its practical relevance, the current study extends research on the adaptive use of decision strategies to the group level and addresses the following questions: Do groups learn to select the decision strategy that fits best to a novel environment, and how well do they do so in comparison to individuals?

Comparing Individuals with Groups

Comparing individual with group performance has a long tradition in psychology (e.g., Watson, 1928) that has documented both the superiority of groups as well as their inferiority to individuals under certain conditions. Some of the inconsistencies can be resolved when taking the specific task context and methodology employed into account, as performance of individuals and groups is a function of the available resources, strategies of their use, task context, and methodology (cf. Bottger & Yetton, 1988; Hill, 1982). We will add to this list that the environmental structure plays an important role too (cf. Gigerenzer et al. 1999). For a fair comparison between individual and group performance, it is additionally important to specify the dependent measure: The performance of a collective (or interactive) group can be compared to (1) the average individual performance, (2) the most competent member of a statistical aggregate or nominal group (Hill, 1982), and / or (3) a statistically pooled response (e.g., averaging continuous guesses in research on the wisdom of crowds, see e.g., Lorenz, Rauhut, Schweitzer, & Helbing, 2011). For example, research shows that collective groups outperform the average individual on intellective tasks, that is, tasks for which a correct answer exists and is demonstrable (for an overview see Kerr & Tindale, 2004). In highly demonstrable tasks, groups are likely to adopt the opinion of the best member ("truth wins"), so that groups may also perform at the level of their best member. Very few studies have shown that groups may even outperform their best members (e.g., Laughlin, Bonner, & Miner, 2002). In brainstorming research, on the other hand, collective groups were shown to underperform nominal groups in terms of quantity of generated ideas (for an overview see Stroebe, Nijstad, & Rietzschel, 2011). In terms of memory capacity, collective groups were shown to remember more than the average individual but less than nominal groups (Betts & Hinsz, 2010). These few examples illustrate that no general conclusion concerning group superiority can be drawn and that the comparison measure matters.

To assess group performance in the following experiments, we will therefore compare it with the average as well as best individual of a nominal group. Besides being a statistical benchmark, nominal groups can be seen as simulating a group decision process, in which members observe each other's performance on the first trials or receive feedback about each other's performance in a similar task, and then agree on following the suggestions of the best member instead of deciding on every trial jointly. If collective groups perform at the level of nominal groups, neither process losses nor process gains had an impact on their performance (Steiner, 1972). Not reaching the potential would point to process losses, for example, due to coordination difficulties (Steiner, 1972), production blocking (Diehl & Stroebe, 1987) or distraction (Baron, 1986).

By studying how well groups learn to use the appropriate strategy in an unknown task environment, we extend research that compares individual with group performance to a strategy learning task. At the same time we add to decision making research that has focused on the ability of individuals to adaptively select strategies in different environments (Bröder, 2003; Rieskamp & Otto, 2006). For example, task characteristics such as costs of information search or time pressure were found to foster limited information search and noncompensatory ways of information integration (e.g., Bröder, 2003; Christensen-Szalanski, 1978, 1980; Payne et al., 1988, 1993). Moreover, environment characteristics such as the dispersion of cue validities and information redundancy have been found to influence decision making in a systematic way (e.g., Dieckmann & Rieskamp; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2006). As groups can be conceptualized as information processing entities where cognition is distributed across individuals (De Dreu, Nijstad, & van Knippenberg, 2008; Hinsz, Tindale, & Vollrath, 1997; Levine & Smith, 2012) and groups face similar conditions like individuals when making decisions, we expect that the same principles found for individuals also hold for groups. Our first hypothesis is therefore that groups are able to learn to use appropriate decision strategies contingent on the task environment. We also ground this prediction on research on group decision making that has shown that groups apply similar decision strategies to those applied by individuals (Reimer, Hoffrage, & Katsikopoulos, 2007; Reimer & Katsikopoulos, 2004). Lastly, we base our prediction on organizational psychological research on team adaptive capacity (i.e., the capacity to gather information from the environment and "to make functional adjustments", cf. Randall et al., 2011, p. 526) that certifies groups *adaptive team performance* when encountering novel conditions in a number of applied settings (such as by airline crews, Waller, 1999; see also Burke et al., 2006; LePine, 2003).

How fast will groups learn to adapt their decision strategy? One important mechanism behind strategy selection is learning from feedback (Rieskamp & Otto, 2006). While feedback generally enhances learning and motivation (Nadler, 1979), studies in psychology (e.g., Davis, 1969; Laughlin & Shippy, 1983; Tindale, 1989; see Hill, 1982, and Hinsz et al., 1997, for reviews) and behavioral economics (Kocher & Sutter, 2005; Maciejovsky, Sutter, Budescu, & Bernau, 2010) have shown that groups require fewer feedback trials than the average individuals to reach asymptotic levels of learning. Reasons for this superiority of groups may be a stronger reliance on memorization (Olsson, Juslin, & Olsson, 2006) and better processing of feedback information (Hinsz, 1990). This leads us to the expectation that groups will learn faster to select adaptive decision strategies than individuals in any environment:

In sum, we hypothesize that groups will learn faster to adapt their decision strategy to an unfamiliar environment over time than the average individual. We will run exploratory analyses to test whether they will even be as good as the best individual.

Two Prototypical Decision Strategies

To investigate these hypotheses, we conducted two experiments on a two-alternative forced choice task, in which participants had to select the more profitable oil-drilling site. Each alternative (i.e., oil-drilling site) was described on a range of attributes (henceforth: cues) such as the results of seismic analysis. In line with research on individuals (e.g., Rieskamp & Otto, 2006), our focus was on environments in which two prototypical decision strategies work well: take-the-best (Gigerenzer & Goldstein, 1999) and weighted additive (WADD). Both strategies make predictions about the information search and choice behavior (Bröder, 2003; Payne et al., 1988; Rieskamp & Otto, 2006), and their success depends on the environmental structure.

Take-the-best looks up the best (i.e., most valid) cue for both alternatives. If this cue discriminates between them (i.e., is positive for one, but negative for the other), take-the-best selects the alternative with the positive cue value and ignores all other cues (Gigerenzer & Goldstein, 1999). Think of our introductory example: if the group considers seismic analysis as the most valid cue and it indicates a high quality for oil-drilling site X but not for Y, the group would administer no further tests and would choose oil-drilling site X. If seismic analysis, however, showed positive results for both sites, a group using take-the-best would acquire the next best cue, and so on, until a discriminating cue is found. A frequent criticism is that people violate the stopping rule and search for more information than necessary, that is, acquire information after the first discriminating cue (Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003). This is particularly common when information search does not incur any costs (e.g., Dieckmann & Rieskamp, 2007). However, others have argued that it does not rule out take-the-best when people look up too many cues as long as the final choice is based on a single cue (see Hogarth & Karelaia, 2007). In this regard, our experiment constitutes a hard test bed as information search did not incur any costs. We will report a method to test whether unnecessarily acquired information influenced the decision, which would more strictly speak against a consistent use of take-the-best than the mere numbe of acquired cues.

In contrast, WADD looks up all cues for both alternatives, multiplies each cue value by its weight, and then selects the alternative with the larger weighted sum. Variants of WADD take instead of the validities chance-corrected validities (Glöckner & Betsch, 2008) or logodds as weights (termed Naïve Bayes; Bergert & Nosofsky, 2007; Lee & Cummins, 2004). Strictly speaking, WADD is assumed to integrate all available cues (e.g., Czerlinski, Gigerenzer, & Goldstein, 1999). However, WADD also works with limited information search, namely, if one assumes that WADD searches cues sequentially according to their validity and stops search as soon as no additional cue can overrule a preliminary decision (cf. Rieskamp & Dieckmann, 2012). On this basis, we can define "necessary information" as the minimum number of cues WADD has to search for so that no additional cue could possibly compensate for the decision based on the acquired cues. Searching for fewer than necessary cues would violate the search rule of WADD (Hogarth & Karelaia, 2007). The advantage of these two models is that they formulate testable predictions on information search, stopping, and choice rules, which can also be tested in groups.

As this is the first study that studies the adaptive use of take-the-best and WADD in groups, we also intend to explore *how* groups apply strategies as compared to individuals. Is the accordance with the strategy's search and stopping rules higher in groups than in individuals? Do groups apply strategies more consistently than individuals (Chalos & Pickard, 1985)? We will explore these questions on the basis of process and outcome data.

Experiment 1

Experiment 1 constitutes a first test bed for our assumptions on adaptive strategy selection in groups as opposed to individuals. To investigate whether participants learn to select strategies adaptively, that is, contingent on the environmental structure, we randomly assigned them to one of two environments, which were constructed to discriminate between the use of take-the-best and WADD: Take-the-best led to the highest performance in the take-the-best-friendly environment and WADD in the WADD-friendly environment. In such environments, people's accordance with the best-performing (i.e., adaptive) strategy has been shown to increase over time when working alone (Bröder, 2003; Bröder & Schiffer, 2006; Rieskamp & Otto, 2006). The task in each case was to select the more profitable of two oil-drilling sites based on a range of cues, with outcome feedback after each trial. Participants were randomly assigned to work alone or in two-person groups (hereafter: dyads).

Method

Participants

Participants included 120 people (60 females; M_{age} = 26.3 years, *SD* = 3.7), of whom 77% indicated being a student. Participants received €12.96 on average (*SD* = 0.83; €1 ≈ $1.37 at the time). To complete the experimental task, individuals took on average 36 min (*SD* = 12) and dyads 50 min (*SD* = 21).

Design and procedure

The experiment had a 2 × 2 × 3 (Participant [individual, dyad] × Environment [take-the-best-friendly, WADD-friendly] × Block) factorial design. The first two factors were between subjects, the third within subject. Upon arrival, participants were randomly assigned to one of the four between-subjects conditions, forcing equal cell sizes of 20 units. They were seated in front of a touch screen either individually or in same-sex dyads. After answering demographic questions, participants completed a practice trial and then worked on the experimental task. Dyads were encouraged to discuss their information search and agree on a joint decision (see Appendix A for instructions).

Experimental task

The oil-drilling task (Czienskowski, 2004) is a MouseLab-like task (Payne et al., 1988) that asks participants to choose the more profitable of two oil-drilling sites in a sequence of trials. Each oil-drilling site was described by six cues and their validities (which correspond to the actual validities in the set; see Figure 1). Validities in decreasing order in both environments were (in percentages, with the discrimination rates for the take-the-best-friendly and WADD-friendly environment in parentheses): 78% (.35; .69), 71% (.54; .65), 65% (.65; .77), 60% (.58; .58), 56% (.69; .69), 53%

(.58; .58).[2] Cues appeared in alphabetical order. Cue validities and cue names were randomly paired once before the experiment and stayed fixed throughout the experiment and for all participants. “Validity” was described as the proportion of correct answers using that cue alone when the cue was applicable (i.e., discriminated between the two alternatives). The cues were framed as tests that could be commissioned (i.e., clicked on) to inform choice. Figure 1 illustrates the two decision strategies, WADD and take-the-best, with screenshots of the task interface. At the beginning of each trial, all boxes contained question marks. They could be clicked on separately to reveal whether the cue had a positive (“+”) or a negative (“-”) value, which remained visible until a choice was made. Clicking on cues was cost free. Outcome feedback followed each trial. For each correct choice, the participant’s account increased by 1,000 petros, a fictitious currency, and equivalent to €0.10.

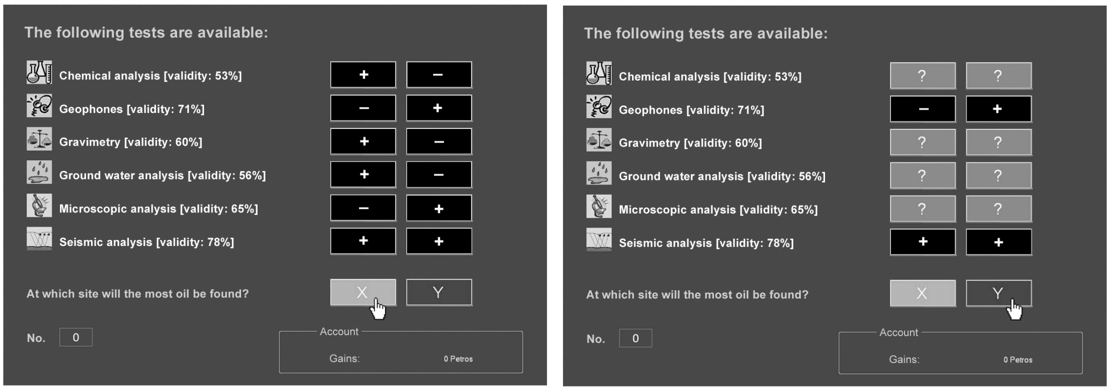


Figure 1. Screenshots of the task interface including six cues for each oil-drilling site (X and Y) illustrating the search behavior of a weighted additive strategy (WADD, left) and take-the-best (right). WADD required looking up all cues to calculate the weighted sum for each alternative. Take-the-best looked up the cue with the highest validity (here: seismic analysis) first, and, as this one did not discriminate, it looked up the cue with the second highest validity (geophones) next. As this cue discriminated, take-the-best reached a decision and ignored the remaining cues, which is why they are still hidden (“?”).

---

[2] The differently high discrimination rate of the most valid cue had no effect on the opening rate of this cue as a a MANOVA with the percentages of trials in which the most valid cue was opened for object A and object B as dependent variables and the two conditions (individuals vs. dyads; take-the-best-friendly vs. WADD-friendly environment) as independent factors revealed (all *F*s < 1).

The task comprised three blocks, each consisting of the same set of two times 26 items (adapted from Rieskamp & Otto, 2006, Study 2; for the complete item sets see Tables A.1 and A.2 in Appendix A). The items within each block were randomly ordered with the restriction that the left and the right oil-drilling sites were equally often correct. Overall, 50% of the total item set were critical items, that is, items for which the two strategies make opposing predictions. To create a WADD-friendly environment, items were constructed by means of genetic algorithms such that WADD reached an accuracy of 88%, while take-the-best reached an accuracy of only 62%. In the take-the-best-friendly environment, accuracies were reversed: 88% for take-the-best and 62% for WADD.[3]

Results

The results section is structured as follows: We first investigate whether participants learned to adapt their strategy to the environment by analyzing performance changes over the three trial blocks. Performance is measured as the percentage of correct trials out of the 156 trials. To better compare performance between individuals and dyads, we will also report analyses on nominal dyads. To evaluate the adaptivity of strategy use, we will focus on accordance rates with the most appropriate strategy in each environment. Lastly, we will test how participants conformed to the corresponding search and stopping rules.

Performance

To investigate performance changes over the three blocks, we conducted an ANOVA with repeated measures with the block as within-subject factor and the environment and individuals vs. dyads as between-subjects factor, and the accuracy per block as dependent variable. Figure 2 depicts the results. Accuracy generally increased over time, F_{block} (1.65, 125.594) = 28.294, $p < .001$, $\eta_p^2 = .27$ (Greenhouse-Geisser

[3] The theoretical accuracy of alternative strategies such as Tally, WADD with chance corrected weights and Naïve Bayes lay in between these two benchmarks. In detail (first value for WADD-friendly environment, second value for TTB-friendly environment), theoretical accuracies were: Tally (.79, .58), WADD with chance corrected weights (.73; .77), Naïve Bayes (.69; .81).

corrected). This improvement was more pronounced in the take-the-best-friendly environment, $F_{\text{Block x Environment}}$ (2, 152) = 15.341, p < .001, η_p^2 = .17. Most importantly, individuals and dyads started from the same level, but dyads then improved more quickly than individuals, $F_{\text{Block x Ind. vs. Dyads}}$ (2, 152) = 4.588, p = .01, η_p^2 = .06. Overall, dyads were not better than the average individual, however, $F_{\text{ind. vs. dyads}}$ (1, 76) = 1.84, p = .18, η_p^2 = .02. Lastly, mean performance was lower in the take-the-best-friendly environment (M_{TTB} = 0.81, SD = 0.05) than in the WADD-friendly environment (M_{WADD} = 0.85, SD = 0.05), $F_{\text{environment}}$ (1, 76) = 11.779, p = .001, η_p^2 = .13.

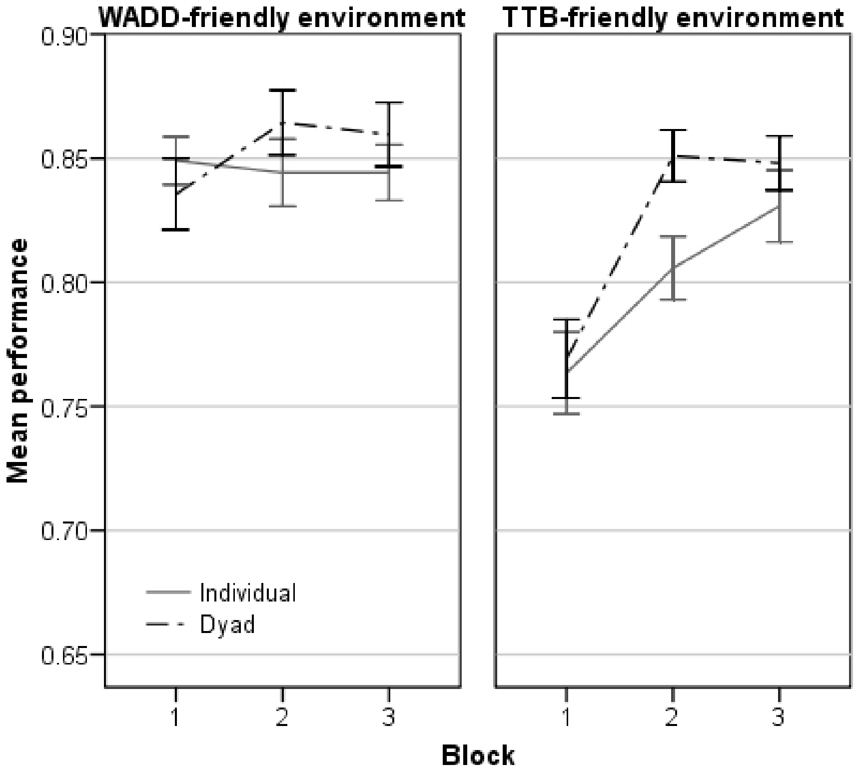


Figure 2. Mean accuracy per block of dyads (n = 20) and individuals (n = 20), in the WADD- (left) and take-the-best- (TTB-) friendly (right) environments. Error bars: ±1 *SE*.

Comparison with the best individual

To create nominal dyads, all 20 individuals of the individual condition in each environment were exhaustively paired, leading to 190 nominal dyads per environment. To determine the performance of each nominal dyad, we took the performance of the "best" (i.e., most accurate) member of a nominal dyad. "Best" was operationalized in two ways: The best individual was either the one who made more accurate choices (a) overall (“best member overall”) or (b) in the first 26 trials, which equals half a block (“best member in 26 trials”). As measure (a) has been criticized for being accessible to the researcher only a posteriori (Miner, 1984), measure (b) is supposed to reflect the idea that groups first determine their best member and afterward adopt this person’s choices (Henry, 1995).

We found that in both environments real dyads (M_{TTB} = 0.82, SD = 0.05; M_{WADD} = 0.85, SD = 0.05) reached the benchmark provided by the nominal dyads, be it by the best member overall (M_{TTB} = 0.83, SD = 0.04; M_{WADD} = 0.87, SD = 0.03) or by the best member in 26 trials (M_{TTB} = 0.82, SD = 0.05; M_{WADD} = 0.86, SD = 0.04), but did not exceed it.[4]

Strategy use

To understand the reasons for the accuracy differences, we next explored the accordance rates with the two best performing strategies take-the-best and WADD in their respective environments. Accordance rates measure how often the strategy predictions match the actual choices and may be interpreted as a measure of consistency of using a certain strategy. Accordance is highly correlated with accuracy but less influenced by randomness. To illustrate, a (100%) consistent use of the most appropriate strategy in each environment would have resulted in an accuracy of only 88%.

Again, we conducted a repeated-measures analysis of variance (ANOVA) to study strategy use over time. The three blocks were entered as within-subject factor, the two environments and individuals vs. dyads as between-subjects factors, and the accordance rate with the adaptive strategy as dependent variable (see Figure B.1 in Appendix B). Mirroring performance, accordance generally increased over time, F_{block} (1.74, 132.40) = 41.530, p < .001, η_p^2 = .35 (Greenhouse–Geisser corrected). This increase was more pronounced in the take-the-best-friendly environment, $F_{\text{Block} \times \text{Environment}}$ (2, 152) = 22.695, p < .001, η_p^2 = .23. Again, dyads adapted more quickly than

[4] We did not test these differences statistically because of the largely unequal sample sizes (n = 190 nominal dyads vs. n = 40 real dyads; cf. Field, 2009). Moreover, it can be seen from the values that no practically relevant differences can be observed.

individuals in the take-the-best-friendly environment and were to a greater extent in accordance with WADD in the last block of the WADD-friendly environment, $F_{\text{Block} \times \text{Individuals vs. Dyads}}$ (2, 152) = 3.284, p = .04, η_p^2 = .04. No overall differences between individuals and dyads were revealed, $F_{\text{individuals vs. dyads}}$ (1, 76) = 2.195, p = .14, η_p^2 = .03.

Information search and stopping rule

As accordance rates have been criticized for being too imprecise to reveal cognitive processes from behavioral data (Bröder & Schiffer, 2003), we will provide in the following some additional measures to validate conclusion that participants improved over time because they learned to use the most appropriate strategy. In particular, we will look at information search behavior and investigate how it accorded with the information search and stopping rules predicted by take-the-best and WADD. Before we can do that, however, we have to determine the decision strategy each individual and dyad most likely used. For this, we used Bröder and Schiffer's (2003) maximum-likelihood method of strategy classification. With this method, the best-fitting model from take-the-best, WADD, Tally and guessing[5] can be determined, whereby the fit is determined in reference to the likelihood of the data given the model (see Bröder & Schiffer, 2003, for details).

In the take-the-best-friendly environment, 13 individuals and 18 dyads were classified as adaptively using take-the-best, while in the WADD-friendly environment 18 individuals and 19 dyads were classified as adaptively using WADD. On the surface, they did not differ in their information search, as these participants searched in both environments on average for 80.7% (SD = 16.3) of the available cues (ANOVA: all Fs < 1.7). The number of cues was more than necessary for take-the-best (on average, 4.46 boxes (SD = 2.01) were opened in addition to the first discriminating cue in the take-the-best-friendly environment), indicating that cost-free cues triggered extensive cue acquisition. This is congruent with previous findings, which showed that introducing a search cost after a learning phase led to a decrease in cue acquisition (i.e.,

[5] Tally is considered as a fourth alternative besides the strategies with the highest expected accuracy in the two respective environments and a baseline guessing model, as it is usually done (e.g., Bröder & Schiffer, 2006). Tally (or Dawes' rule; Dawes, 1979) assumes that people sum up the positive cues and choose the option with the larger total sum. It thus searches for all information. In the WADD-friendly environments, it performed second best (79%) and in the take-the-best-friendly environment it performed worse than take-the-best (58%).

an increase in accordance with the take-the-best stopping rule) in a take-the-best-friendly environment but not in a take-the-best-unfriendly environment. This indicated that people had learned different choice rules though not differing in their stopping rule in the learning phase (Dieckmann & Rieskamp, 2007; Rieskamp & Dieckmann, 2012). In fact, searching for cues does not necessarily imply that the cues are integrated; search is often continued to enhance confidence in decisions already made (Harvey & Bolger, 2001; Newell et al., 2003; Svenson, 1992).

We therefore now introduce two more fine-grained measures of strategy use: (1) To validate WADD as a choice rule, we checked how often participants that were classified as adaptive WADD-users opened fewer cues than necessary, in short “too few” (recall that necessary means that no further evidence would overrule the decision based on the acquired cues). (2) To validate take-the-best as choice rule, we analyzed those cases, in which participants who were classified as adaptively using take-the-best opened less valid cues that contradicted the first discriminating (more valid) cue, and checked whether this less valid cue overruled their decision—which, according to take-the-best, it should not. In other words, we counted how often the decision of take-the-best users was overruled by compensatory evidence (“compensatory choices”).

Figure 3 depicts the results for these two measures. In the left panel, the results concerning the WADD-users can be seen. It shows that, in the first block, WADD-users opened on average about 1 box less than necessary. This number decreased over blocks to a mean number of 0.46 boxes, F_{block} (1.458, 48.101) = 11.171, $p < .001$, $\eta_p^2 = .25$ (Greenhouse-Geisser corrected), with no differences between individuals and dyads, $F_{\text{individuals vs. dyads}}$ (1, 32) = 2.373, $p = .13$, $\eta_p^2 = .07$. In other words, WADD-users became more consistent with their search rule but still opened slightly too few boxes.

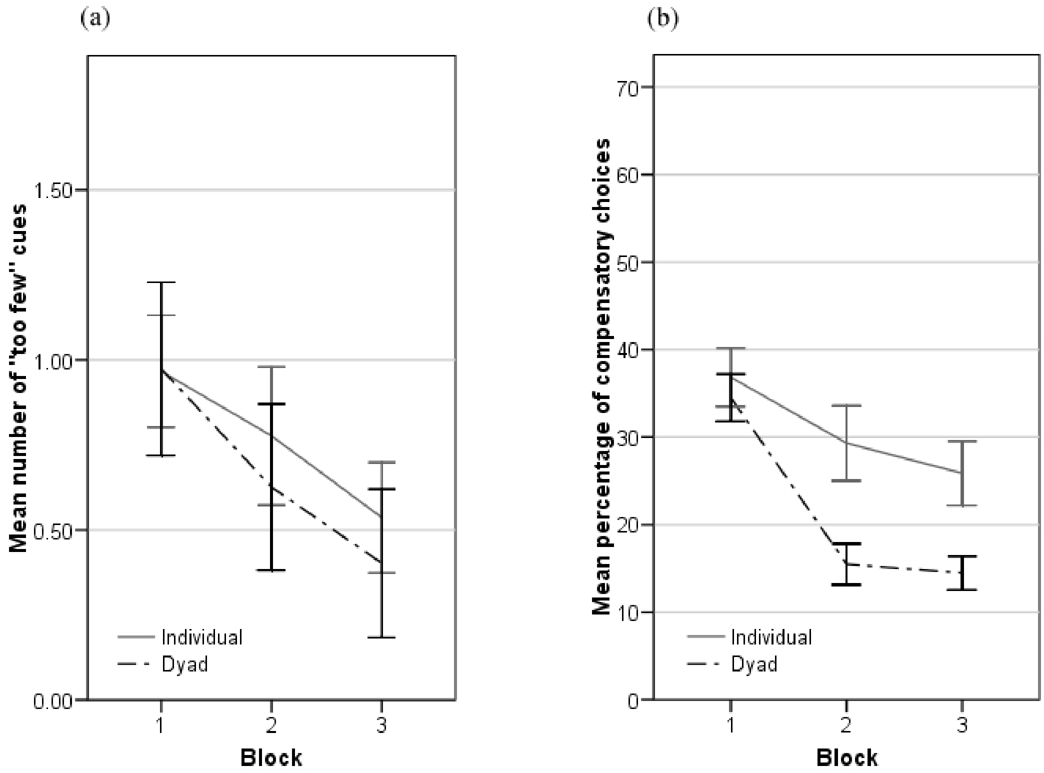


Figure 3. Two measures of strategy use concerning the stopping rule, (a) in the WADD-friendly environment and (b) in the take-the-best friendly environment. (a) The left panel depicts the mean number of too few cues that have been looked up, that is, cues that should have been opened so that the decision could not be overruled by additional evidence. This measure was calculated for the 18 individuals and 19 dyads who were classified as adaptive WADD-users. (b) The right panel depicts the proportion of those trials in which people decided against the first discriminating cue based on less valid cues that were additionally opened, although, according to take-the-best, these less valid cues should not have overruled the first discriminating cue. This measure was calculated for the 13 individuals and 18 dyads who were classified as adaptive take-the-best-users. Error bars: ±1 *SE*.

In the right panel of Figure 3, the results concerning the adaptive take-the-best users can be seen. It shows the percentage of those cases, in which participants saw contradictory evidence,[6] which overruled the decision as suggested by take-the-best. In

[6] The amount of contradictory evidence can be measured in different ways, for example, by calculating the weighted sum of all those cues that were opened after the first discriminating one, for each option X and Y, and compare these sums with each other. If the first discriminating cue is pointing to X, for example (i.e., having a positive value for X), but the weighted sum of cues opened after the first discriminating one is larger for Y, this is regarded

the first block, individuals and dyads decided in around 35% of cases, in which they saw contradictory evidence, against take-the-best. Over time, this proportion decreased, F_{block} (2, 52) = 29.909, $p < .001$, $\eta_p^2 = .54$, and it did so more strongly for dyads (where it decreased to about 15%) than for individuals (where it decreased to about 25%), $F_{\text{Block} \times \text{Ind. vs. Dyads}}$ (2, 52) = 3.654, $p = .03$, $\eta_p^2 = .12$.

Summary

To summarize, in Experiment 1 we sought to test whether and how well individuals and dyads learned to select the appropriate strategy in an unknown task environment. It provided some evidence that not only individuals but also dyads are able to adapt to different, but stable environmental structures. Dyads even showed a faster adaptation process, but they did not surpass the potential given by the best individual. The high performance rates were supported by the finding that the majority of participants were classified as using the adaptive strategy. Accordance rates mirrored performance results and indicated a more consistent use of take-the-best by dyads. Convergent evidence came from process measures: information search became more consistent over time, and again to a greater extend for dyads in the take-the-best friendly environment.

Experiment 2

In Experiment 2 we sought to replicate the findings of Experiment 1 and extend them to a task in which environmental structures changed over time and a new strategy had to be learned. Experiment 2 thus comprised two phases: the learning phase, which was identical to Experiment 1 and varied the environmental structure between participants, and the relearning phase, in which participants were confronted with the alternative environment. Consequently, each participant encountered both environments (the take-the-best-friendly and the WADD-friendly) from Experiment 1,

---

as a trial with contradicting evidence. We report the results for this measure. An alternative way would be to count the number of discriminating cues that follow the first discriminating one and to which direction they point. If, after the first discriminating cue, more discriminating cues follow that point into the other direction (Y), this would be regarded as contradicting evidence. These measures yield very similar results.

one after the other. Experiment 2 thus provides a stricter test for adaptive strategy selection by varying the environmental structure within subjects, as Payne et al. (1988) have suggested.

Since Experiment 2 contained a change in the environment that rendered another strategy adaptive, it differed in some important aspects from Experiment 1. While the learning phase of Experiment 2 was equivalent to Experiment 1 (with the difference that people were informed at the beginning that there would be two phases), the relearning phase of Experiment 2, though structurally corresponding to the learning phase, required additional subtasks. These subtasks were (a) to detect the need for change, (b) to find and apply a new and better strategy than the one selected in the learning phase, and (c) to overcome a—now maladaptive—routine established in the learning phase.

When people are faced with familiar problems, routinized decision behavior has many advantages, such as allowing for efficiently dealing with a situation and for immediately reacting and performing well. On the group level, having developed a routine reduces the need for consideration, coordination, and negotiation (Gersick & Hackman, 1990). When a situation changes, however, and some novel decision behavior is—unnoticeably—required, routines become maladaptive. In fact, individuals as well as groups have difficulty overcoming maladaptive routines, especially with increasing routine strength or when they are under time pressure (e.g., Betsch, Fiedler, & Brinkmann, 1998; Betsch, Haberstroh, Glöckner, Haar, & Fiedler, 2001; Bröder & Schiffer, 2006; Reimer, Bornstein, & Opwis, 2005; for a review of theories, see Betsch, Haberstroh, & Höhle, 2002). The additional requirements make the relearning phase more difficult than the learning phase of Experiment 2 and than Experiment 1. We thus expected an overall lower performance in the relearning phase. This enhanced difficulty has one additional advantage, as it leaves more room for learning to take place. In fact, one could argue that in Experiment 1, the lack of learning in the WADD-friendly environment was observed due to a ceiling effect as participants, both individuals and dyads, had started out with an already very high accordance to WADD. If performance is already high and people do the upper benchmark of performance, they might not see any need to change their strategy, which might be the reason for the lack of further improvement in the WADD-friendly environment.

Methods

Participants

Participants included 120 people (60 females; M_{age} = 24.2 years, *SD* = 3.7), of whom 83% indicated being a student. Participants received €24.7 on average (*SD* = 1.55). To complete the oil-drilling task, individuals took on average 53 min (*SD* = 15) and dyads 72 min (*SD* = 24).

Design and procedure

Again, the experiment had a 2 × 2 × 3 (Participant [individual, dyad] × Starting Environment [take-the-best friendly, WADD friendly] × Block) factorial design, and phase as an additional factor (Phase 1, Phase 2). The first two factors were between subjects, the third and fourth within subject. Upon arrival, participants were randomly assigned to one of the four between-subjects conditions, forcing equal cell sizes of 20 units. As in Experiment 1, participants worked with a touch screen either individually or in same-sex dyads. After answering demographic questions, participants completed a practice trial and then worked on the experimental task, which was exactly the same in each phase as in Experiment 1. The difference was that this time all participants worked on the two environments consecutively, one half first on the take-the-best-friendly environment and then on the WADD-friendly environment with a break in between, the other half on the reverse order. Participants were told at the very beginning that they had to work on two phases, finding profitable oil-drilling sites first in the United States and then in Argentina (or vice versa, counter-balanced per environment). We provided this country hint in all conditions to suggest to participants that something might have changed and to thereby secure a minimum level of adaptivity; it has previously been shown that without a hint almost no adaptivity can be observed in a changing environment, resulting in a floor effect (Bröder & Schiffer, 2006).

Results

Performance

To investigate performance changes over the three blocks of each phase, the percentage of correct trials was entered into a repeated-measures ANOVA with the three blocks and the two phases as within-subject factors, and the starting environment and individuals vs. dyads as independent variables. As can be seen in Figure 4, accuracy generally increased over time in both phases, F_{block} (1.82, 138.57) = 90.458, $p < .001$, $\eta_p^2 = .54$ (Greenhouse–Geisser corrected). This increase was more pronounced in the take-the-best-friendly environment, independent of the phase, $F_{\text{Block} \times \text{Environment}}$ (2, 152) = 2.929, $p = .06$, $\eta_p^2 = .04$. Dyads were on average better than individuals, $F_{\text{ind. vs. dyads}}$ (1, 76) = 3.939, $p = .05$, $\eta_p^2 = .05$. This difference was driven by the take-the-best-friendly environment: dyads were better in the take-the-best-friendly environment in both the learning and the relearning phase (though to a lesser degree in the second phase), but did not differ from individuals in the WADD-friendly environment in both phases, $F_{\text{Phase x Ind. vs. Dyads} \times \text{Environment}}$ (1, 76) = 3.601, $p = .06$, $\eta_p^2 = .05$. Moreover, different learning curves were observable: individuals mainly improved from the first to the second block, but dyads kept on improving to reach a higher final level, $F_{\text{Block} \times \text{Ind. vs. Dyads}}$ (2, 152) = 3.617, $p = .03$, $\eta_p^2 = .05$.

As expected, average performance of all participants dropped from the first to the second phase, F_{phase} (1, 76) = 63.416, $p < .001$, $\eta_p^2 = .46$. In other words, participants suffered from the change in the environment. However, the direction of change played an important role. Learning to apply WADD in the second (relearning) phase when it had not been adaptive before was more likely than adopting take-the-best as a novel strategy. Thus, when the take-the-best-friendly environment constituted the starting environment, participants' performance did not differ between the phases. This was not the case in the reverse experimental order. In both phases, performance was higher in the WADD-friendly environment than in the take-the-best-friendly environment, respectively. The drop from the first to the second phase was much less pronounced when the WADD-friendly environment constituted the second environment than when the take-the-best-friendly environment came second, $F_{\text{Phase x Environment}}$ (1, 76) = 52.855, $p < .001$, $\eta_p^2 = .41$, which indicated a preference for WADD.

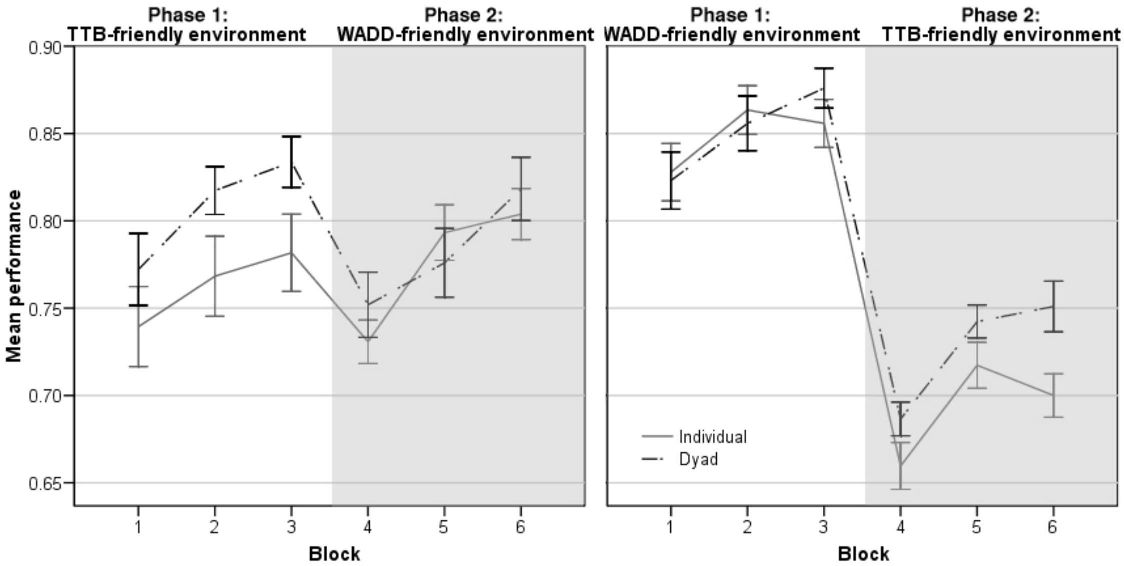


Figure 4. Individuals' and dyads' average performance in the two experimental orders: (a) The left panel depicts the rates of performance with the adaptive strategies in the experimental order of first the take-the-best-friendly and then the WADD-friendly environment; (b) the right panel depict the results for the reverse order. Error bars: ±1 *SE*.

Comparison with the best individual

Again we compared the performance of real dyads with that of nominal dyads. Nominal dyads were composed by exhaustively pairing the 20 individuals of the individual condition of each environment and performance was determined by giving each nominal dyad the score obtained by the better of the two individuals ("best member overall" and "best member in 26 trials"). In the take-the-best-friendly environments, real dyads ($M_{\text{phase 1}} = 0.81$, $SD = 0.07$; $M_{\text{phase 2}} = 0.73$, $SD = 0.04$) reached the baseline provided by the nominal dyads in both phases, be it by the best member overall ($M_{\text{phase 1}} = 0.82$, $SD = 0.05$; $M_{\text{phase 2}} = 0.73$, $SD = 0.03$) or the best member in the first 26 trials ($M_{\text{phase 1}} = 0.81$, $SD = 0.05$; $M_{\text{phase 1}} = 0.71$, $SD = 0.04$). Also in the WADD-friendly environments, real dyads ($M_{\text{phase 1}} = 0.85$, $SD = 0.06$; $M_{\text{phase 2}} = 0.78$, $SD = 0.08$) were close to the performance of the best member overall ($M_{\text{phase 1}} = 0.88$, $SD = 0.03$; $M_{\text{phase 2}} = 0.81$, $SD = 0.04$) and of the best member in 26 trials ($M_{\text{phase 1}} = 0.87$, $SD = 0.03$; $M_{\text{phase 1}} = 0.79$, $SD = 0.05$).

Strategy use

Strategy use over time (i.e., accordance rate of the adaptive strategy in each environment) was entered into a repeated-measures ANOVA with the three blocks and two phases as within-subject factors and the environmental of the first phase and individuals vs. dyads as independent variables (see Figure B.2 in Appendix B).

Within each phase, accordance generally increased over time, F_{block} (1.693, 128.705) = 119.992, $p < .001$, $\eta_p^2 = .61$ (Greenhouse–Geisser corrected). Like performance, average accordance with the adaptive strategy dropped from the first phase to the second, F_{phase} (1, 76) = 100.145, $p < .001$, $\eta_p^2 = .57$; this drop was particularly deep when participants were confronted with the take-the-best friendly environment in the second phase, $F_{\text{Phase x Environment}}$ (1, 76) = 28.770, $p < .001$, $\eta_p^2 = .28$; and increase in accordance was steepest in this environment and phase too, $F_{\text{Block x Phase x Environment}}$ (2, 152) = 12.594, $p < .001$, $\eta_p^2 = .14$. Overall, accordance with the adaptive strategy was lower in the take-the-best-friendly environment than in the WADD-friendly environment, $F_{\text{environment}}$ (1, 76) = 7.132, $p = .01$, $\eta_p^2 = .09$.

Dyads achieved higher accordance rates with take-the-best in the take-the-best-friendly environment than individuals in both phases, but slightly lower accordance rates with WADD in the WADD-friendly environment in both phases, $F_{\text{Phase x Ind. vs. Dyads} \times \text{Environment}}$ (1, 76) = 8.201, $p = .01$, $\eta_p^2 = .10$, so that dyads only slightly surpassed individuals in overall accordance with the most adaptive strategy ($M_{\text{individuals}} = .77$, $SD = .01$ vs. $M_{\text{dyads}} = .80$, $SD = .01$), $F_{\text{ind. vs. dyads}}$ (1, 76) = 3.454, $p = .07$, $\eta_p^2 = .04$.

Information search and stopping rule

Again we used the maximum-likelihood method of Bröder and Schiffer (2003) to classify participants as using one of the following strategies: take-the-best, WADD, Tally, or guessing. In the first phase, 16 individuals and 18 dyads were classified as adaptively using take-the-best in the take-the-best-friendly environment. In the WADD-friendly environment, 18 individuals and 18 dyads were classified as using WADD. In the second phase, no individual and only seven dyads were classified as adaptively using take-the-best in the take-the-best-friendly environment. In the WADD-friendly environment, more participants, namely thirteen individuals and thirteen dyads, were classified as adaptively using WADD, probably indicating that WADD was either easier to learn or a default strategy when encountering a changing environment, as other studies have argued before (e.g., Bröder & Schiffer, 2006).

Restricting the number of participants to the adaptively classified and entering individuals vs. dyads and the environment as independent variables and the mean number of acquired cue as dependent variable into an ANOVA per phase revealed that participants in the first phase searched for more information in the WADD-friendly environment (M = 84.3%, SD = 14.0) than in the take-the-best-friendly environment, where search was still quite high (M = 69.1%, SD = 20.4), $F_{\text{environment}}$ (1, 66) = 12.899, p = .001, η_p^2 = .16. Due to the lack of classified individuals as take-the-best users in the second phase only a comparison within dyads was possible. Here, the mean number of acquired cues was not an indicator of strategy use as no differences were revealed between environments (overall M = 77.8%, SD = 14.3). This amount of information acquisition again exceeded the amount required by take-the-best (on average, 3.75 boxes (SD = 2.12) were opened after the first discriminating cue in the first phase and 6.59 boxes (SD = 1.72) in the second phase in the take-the-best-friendly environment).

We next analyzed how many fewer cues than necessary were opened by the adaptive WADD-users. Figure 5a depicts the results for the first half. An ANOVA with repeated measures revealed that individuals and dyads became more consistent with the WADD stopping rule over time, F_{block} (1.515, 51.527) = 10.795, p < .001, η_p^2 = .24 (Greenhouse-Geisser corrected). In the second half, participants started with opening on average 1.30 cues (SD = 1.10) too few, which decreased to 0.90 (SD = 1.15) in the last block, again indicating an increasing consistency with WADD though the absolute numbers were higher than in the first phase, F_{block} (1.136, 27.276) = 3.987, p = .03, η_p^2 = .14 (Greenhouse-Geisser corrected; see Figure 5b).

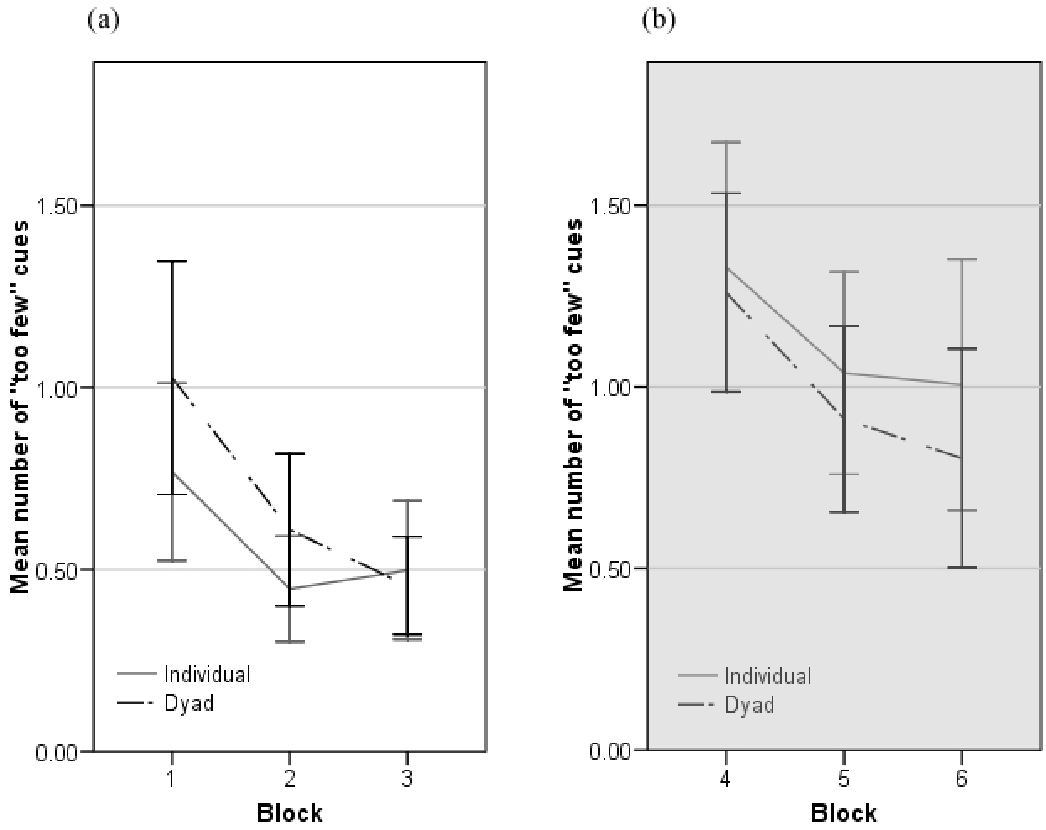


Figure 5. Mean number of "too few" cues opened by participants who were classified as WADD-users in the WADD-friendly environment, (a) in the first phase (n = 18 individuals and n = 18 dyads) and (b) in the second phase (n = 13 individuals and n = 13 dyads). Error bars: ±1 *SE*.

Figure 6 depicts the proportion of trials, in which adaptive take-the-best users, who opened more contradictory than supportive evidence after the first discriminating cue, were influenced by this evidence and chose the option not favored by the first discriminating cue. As in experiment 1, a steady decrease in those compensatory choices is observable also in the first phase, F_{block} (2, 60) = 26.985, p < .001, η_p^2 = .47. In phase 2, no comparison between individuals and dyads was possible as only seven dyads but no individuals were classified as adaptive take-the-best users. They showed a similar decreasing trend, though on a higher absolute level, F_{block} (2, 12) = 39.148, p < .001, η_p^2 = .87.

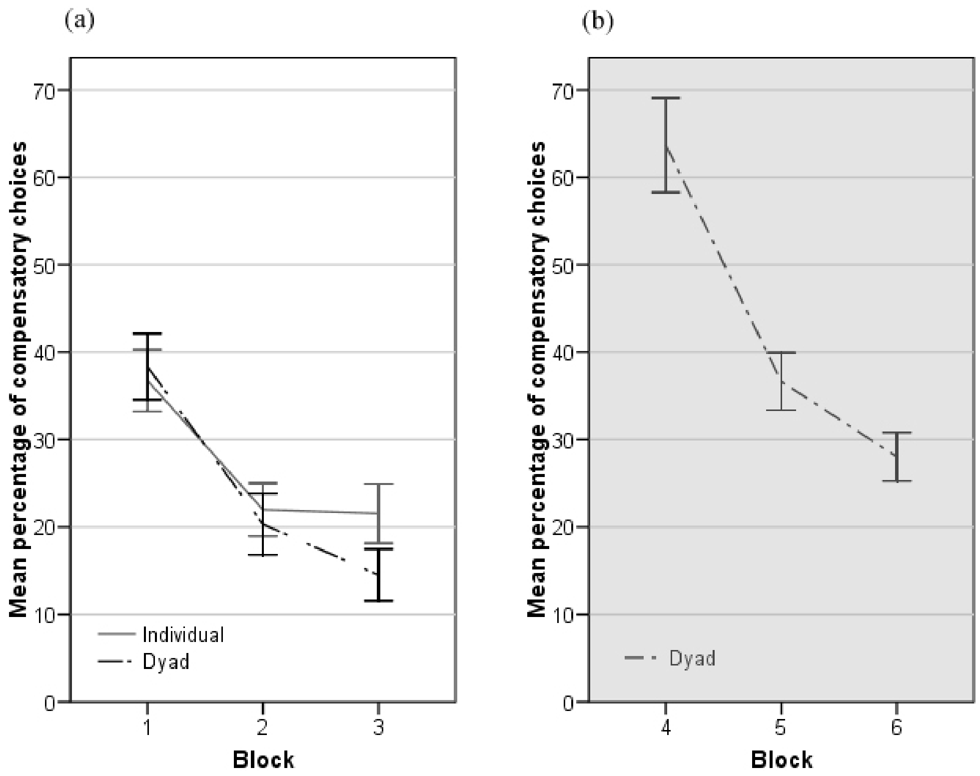


Figure 6. Average proportion of compensatory choices (i.e., deciding against the first discriminating cue) out of trials in which more contradicting evidence than supporting evidence was opened, (a) in the first phase (n = 16 individuals and n = 18 dyads) and (b) in the second phase (n = 7 dyads). This measure was calculated for those participants who were classified as adaptive take-the-best-users. Note that no individuals were classified as take-the-best users in the second phase, why no results can be displayed for individuals in the right panel. Error bars: ±1 *SE*.

Summary

In sum, Experiment 2 mainly replicated the findings of Experiment 1 and tested them in a relearning phase. In the learning phase, dyads were superior to individuals in learning to adaptively follow take-the-best but did not differ in following WADD. The relearning phase apparently constituted a much harder test bed, showing that performance was much lower than in the learning phase. Again, dyads were superior to individuals in learning to adaptively follow take-the-best but did not differ in following WADD. Dyads performed at the level of the best members. Strategies were more consistently used in the first phase than in the second, and dyads applied take-the-best more consistently than individuals, which was indicated by information search behavior and accordance rates and more extremely shown by the classification, which revealed that no single individual was using take-the-best in the second phase.

Discussion

Two experiments were conducted to investigate whether and how well two-person groups, or dyads, as opposed to individuals, adaptively follow decision strategies (here: the prototypical strategies take-the-best and WADD). While Experiment 1 tested adaptive strategy selection in a between-subjects design by allocating participants to either a take-the-best-friendly or a WADD-friendly environment, Experiment 2 employed a within-subject design, in which all participants faced both environments, one after the other. This rendered Experiment 2 a stricter test of adaptive strategy selection, because a new strategy became adaptive and had to be learned. In fact, abandoning a maladaptive routine in the second phase became a very hard task for most participants. Besides replicating previous findings of adaptive strategy selection in individuals (e.g., Bröder, 2003), both experiments provide experimental evidence for adaptive strategy selection in groups. Interestingly, performance analyzes revealed that groups were better than the average individual at learning take-the-best when it was adaptive, that is, required fewer trials to reach asymptotic performance levels (Experiment 1) or achieved on average higher performance levels (Experiment 2). They did not differ from the average individual, however, when following WADD was adaptive. Comparisons with nominal groups revealed that groups performed at the level of the best individuals.

Performance Differences

We hypothesized that groups would be able to adapt to an unfamiliar environment by following the most appropriate strategy. This hypothesis is based on the assumption that groups can be conceptualized as information processing systems and hence apply similar (cognitive) strategies like individuals (Hinsz et al., 1997). We found that groups were as good as the average individual in experiment 1 and better in experiment 2, which was mainly driven by performance differences in the take-the-best-friendly environment.

One reason for this high group performance might be that groups had a higher probability of containing at least one individual who was above the mean ability level of people working alone (Lorge & Solomon, 1955). We therefore compared performance levels of the interacting groups with that of nominal groups. If real groups perform better than the best member of a nominal group, this would be an indicator of process gains from interaction. If real groups perform worse than nominal groups and thus do not reach their potential, process losses have occurred (Steiner, 1972). In our case, we found that real groups performed by and large as well as nominal groups in both environments. In other words, they were not much better than had they only identified the best individual. This finding could be taken to argue against investing in group interaction. Some caution is warranted, though, because another conclusion could be that it is already satisfactory for groups to reach the potential given by the performance of their best member, since groups rarely perform better than individuals, according to a vast amount of literature (e.g., Kerr & Tindale, 2004; Laughlin et al., 2002; Tindale & Sheffey, 2002). Moreover, groups usually have difficulties with identifying their best member without help (e.g., Henry, 1995; Henry, Strickland, Yorges, & Ladd, 1996). Even more relevant may be that group decision making has other advantages, such as legitimacy and acceptance, which may play an important role in many organizational contexts (see Allen & Hecht, 2004, for more benefits).

Much more pronounced than overall performance differences between individuals and groups were performance differences between environments. Performance was higher in the WADD-friendly environment than in the take-the-best-friendly environment, indicating that WADD was easier to learn. This asymmetry in favor of WADD replicates previous findings on the individual level (Bröder, 2003; Bröder & Schiffer, 2003, 2006; Rieskamp & Otto, 2006) and extends them to the group level. Replicating this asymmetry for the relearning phase as well, that is, showing that learning WADD is more likely than learning take-the-best even when it has not been adaptive before, supports the idea of WADD being a default strategy. The following reasons for an asymmetric preference for WADD are discussed in the literature: It may stem from the belief that "more is better" (Chu & Spires, 2003) or simply reflect an exploration phase in which people try to get a sense of which pieces of information are useful before settling on a decision strategy (McAllister, Mitchell, & Beach, 1979). Similarly, Hogarth and Karelaia (2006) argued from a prescriptive perspective that in unknown environments linear models perform better than one-reason decision strategies. From a descriptive perspective, it may also reflect an overgeneralization of the applicability of normally reasonable strategies (Payne et al., 1993, p. 206) and may have been enhanced by leading the participants to focus on accuracy, which has been found to foster WADD (Creyer, Bettman, & Payne, 1990).

What this result also indicates is that the take-the-best environment constituted a much harder environment, and most differences between individuals and groups were found here (as discussed below).

Learning Curves

Previous research has found that groups are faster in learning (e.g., Davis, 1969), and our results point into the same direction. Again, the environment played an important moderator: The learning curve in the take-the-best-friendly environment was steeper for dyads than for individuals, with either individuals reaching the same level of accordance in the final block (Experiment 1) or dyads staying on a higher level in all blocks (Experiment 2). In the relearning phase of Experiment 2, routine effects led to an overall decrease in participants' performance, but mostly when the take-the-best-friendly environment was encountered second. Such negative transfer effects have been widely documented before (e.g., Betsch & Haberstroh, 2005). But although individuals and dyads started at a similarly low level of accordance with the adaptive strategy in Phase 2, the dyads' superiority again became apparent: Dyads were more likely to abandon WADD when it was no longer adaptive, while only the best individuals were successful in doing the same, as the comparison with nominal dyads suggests. In fact, not a single individual was classified as adaptively using take-the-best in the second phase, but seven dyads were. This result also suggests that giving people many opportunities to encounter a novel task that requires abandoning a routine is especially beneficial for dyads, although they might appear as or even more prone to routines than individuals in the first place (Reimer et al., 2005).

Information Search Behavior and Adherence to the Stopping Rule

In a last step, we analyzed consistency of strategy use by measuring accordance with the information search, stopping and choice rule of the respectively most appropriate strategy. All measures indicate an increase in consistency over time and again a higher consistency by groups in the take-the-best-friendly environment. This higher decision consistency in groups is consistent with work by Chalos and Pickard (1985).

As a measure of accordance with the information search and stopping rules, the mean number of acquired cues is usually taken (e.g., Newell & Shanks, 2003; Rieskamp & Dieckmann, 2012). This measure, however, did not allow for separating strategies in our experiments as people opened most of the cues on average (in experiment 1 and the second phase of experiment 2). This is not an uncommon finding in tasks in which information search is cost free (e.g., Bröder, 2003). Though violating in the strict sense the stopping rule of take-the-best in particular, it is not the only indicator of consistent strategy use (cf. Dieckmann & Rieskamp, 2007; Hogarth & Karelaia, 2007). We therefore proposed two more fine-grained measures: the number of cues that were opened too few in order to allow the decision proposed by WADD not to be overruled by subsequent information. As a measure of take-the-best, we analyzed the number of trials in which people were influenced by less valid information. Despite the plausibility of these measures and their insights into strategy use, only a restricted evaluation is possible as no established thresholds exist and no comparison with previous studies is possible. Future studies should further validate these measures.

Summary and Open Questions

To summarize the results, individuals and groups were equally good at applying the default strategy of weighting and adding all available information when this was required (WADD), but groups were better at learning to consider cues in a sequential manner, to ignore irrelevant cues, and to integrate them in a non-compensatory way when it was adaptive (take-the-best). Such a superiority of small groups has been documented before in other learning tasks (e.g., Hill, 1982). This study demonstrates it in a strategy selection task and thus contributes to research on the adaptive capacity of teams (e.g., Burke et al., 2006; Randall et al., 2011).

Plausible explanations for the superiority of dyads in the take-the-best-friendly environment can be derived from the literature that discusses reasons for the superiority of groups in intellective tasks in general (e.g., Laughlin, VanderStoep, & Hollingshead, 1991) and a faster learning rate of groups in particular. These are (a) the greater likelihood of recognizing the correct answer due to a larger sample size; (b) a better joint memory due to better error correction ability (e.g., Hinsz, 1990; Vollrath, Sheppard, Hinsz, & Davis, 1989) and/or better encoding (Weldon, Blair, & Huebsch, 2000; for an overview of findings on collaborative group memory, see Betts & Hinsz, 2010); and (c) the capacity to process more information and use decision rules more consistently (Chalos & Pickard, 1985). Additionally, articulating the decision procedure during discussion may enhance awareness and a deeper processing rendering more likely to detect the appropriate strategy (Kerr, MacCoun, & Kramer, 1996).

The aforementioned reasons, however, would also suggest a superiority of dyads in the WADD-friendly environment, which we did not find. One might argue that a ceiling effect was responsible for not finding this. This explanation, however, can be excluded by taking into account the second phase of experiment 2 where performance dropped and no similarly high levels as in the first phase were reached.

What might explain the asymmetric finding of group superiority when learning take-the-best? We speculate that the possibility for social validation in a dyadic setting may be one reason for our finding that dyads were more prone to be less influenced by irrelevant cues (i.e., cues that were less valid than the best discriminating cue). The approval of one's partner may replace looking up or taking into consideration more cues to feel reassured in one's decision. Another reason may be that a better calibration of cue orderings may be the result of collaborating with another person, as exchanging information with others can speed up learning the order in which cues should be considered (Garcia-Retamero, Takezawa, & Gigerenzer, 2009). Since this was only helpful in the take-the-best-friendly environment, the observed asymmetry may have appeared. It may also be the case that groups per se rather overweight apparently important cues (Gigone & Hastie, 1997), which may be unhelpful in certain environments, such as one that is WADD friendly, but advantageous in others, such as a take-the-best-friendly environment. Last, the information search steps and integration rule of take-the-best might be much easier to verbalize than those of WADD, rendering take-the-best easier to communicate and teach to another person once it had been detected as the appropriate rule (for a related argument that simple, sequential strategies are easier to learn than strategies that weight and add all pieces of information, see Wegwarth, Gaissmaier, & Gigerenzer, 2009). Taken together, the aforementioned aspects might offer some explanations of the superiority imbalance. More research is needed to test these assumptions and shed light on the mechanisms underlying the one-sided superiority.

Moreover, future research should address the question of whether and to what extent this superiority effect can be found "in the wild," that is, in real groups where ignoring irrelevant information may facilitate and improve decision making in certain environments. Admittedly this is an unusual endeavor in the light of much group research that aims at finding ways of fostering the quantity of information considered in and by groups (e.g., Frey, Schulz-Hardt, & Stahlberg, 1996; Larson, Foster-Fishman, & Keys, 1994; Parks & Cowlin, 1996; Stasser, Taylor, & Hanna, 1989; Wittenbaum & Stasser, 1996). This line of research has been stimulated by the repeated finding that groups would not exhaust their potential of pooling more information but mainly discuss shared information known to every member (e.g., Stasser, 1992; Wittenbaum & Stasser, 1996). In those studies, the option with the highest overall sum score was often defined as the best solution though (i.e., Tally; cf. Reimer & Hoffrage, 2012; for a critique see Reimer & Hoffrage, 2006). Therefore, groups ignoring part of the available information necessarily performed worse than the benchmark strategy. This limitation to one type of environment restricts the possible findings concerning group adaptivity. Our results draw an optimistic picture that groups are able to adapt to different environments. The lesson here is that not the mere quantity of information determines the success of a group (cf. Reimer & Hoffrage, 2006) but rather its adaptive integration of information, which may, in certain environments, mean to ignore irrelevant information.

This study is just one step towards studying adaptive strategy selection in groups. Limited generalizability is given by its focus on inferences from givens and a rather abstract, completely unfamiliar experimental task. In everyday life, people may probably experience some resemblance between new and old situations and thus be able to exploit their repertoire of strategies better (without having to encounter the same decision problem for 156 times). A parallel, real-world example may illustrate some of the factors that are missing due to our choice of an unfamiliar task: Take a selection committee where group members receive information on the applicants and meet in order to make a selection. Here, strategic interests may come into play and may influence information sharing and weighting. The example also alludes to factors that may cause more performance differences between individuals and groups in information-intense environments: having to actively search for, remember information and decide which to search for in the first place. The MouseLab-like experimental setup, which presents all available information to participants, certainly simplifies the task in these respects (cf. Glöckner & Betsch, 2008). Therefore, studying more naturalistic settings would be worth to study in future

Another limitation results from our focus on just two environments and two strategies. Although take-the-best and WADD have been identified as two prototypical decision strategies (Bröder, 2003; Rieskamp & Otto, 2006), many more heuristics are part of the toolbox (for an overview see table A.1-1 in Todd & Gigerenzer, 2012, pp.8-9) and many more factors shape the environment than just the payoff structure (such as the time or costs for acquiring information). Therefore, future research may extend our findings to a broader set of decision domains. On the side of the decision maker, further influencing factors, worthy to study, may be intelligence, working memory load (cf. Bröder, 2003), the size of the group, or its composition (Kämmer et al., 2012).

Conclusion

Adaptive capacity is essential for individuals and groups who are engaged in judgment and decision making (Burke et al., 2006; Gigerenzer et al., 1999; Randall et al., 2011). It enables people to adjust their operations in (changing) environments accordingly. The selection of an appropriate strategy from the adaptive toolbox, for example, will lead to efficient and effective decision making in an uncertain environment. Specifically, using take-the-best or WADD when appropriate allows for tailoring one's amount of information needed and way of integrating it to the structure of the environment. The current study showed that both individuals and dyads engage in adaptive strategy selection and that dyads are superior in the adaptive use of take-the-best, but not in the adaptive use of WADD. Thus, in contrast to the common (and partly justified, see Richter, Dawson, & West, 2011) belief of organizations in the superiority of teams (Allen & Hecht, 2004), no generalized verdict in favor of groups can be derived from this study. Instead, it demonstrates how important it is to take the environmental structure of the task into account when comparing individual with group strategy learning and performance.

Acknowledgments

This research was funded by the Max Planck Institute for Human Development, Berlin, Germany. We would like to thank Ulrich Klocke and Torsten Reimer for helpful discussions, Henrik Olsson and two anonymous reviewers for insightful comments on an earlier version of this article, and Anita Todd and Katherine McMahon for editing the manuscript. Thanks are also due to Gregor Caregnato and Jann Wäscher for collecting the data.

Chapter 3

The adaptive use of recognition in group decision making[1]

Kämmer, J. E., Gaissmaier, W., Reimer, T. & Schermuly, C. C.

Abstract

Applying the framework of ecological rationality, the authors study the adaptivity of group decision making. In detail, they investigate whether groups apply decision strategies conditional on their composition regarding task-relevant features. The focus here is on the recognition heuristic, so that task-relevant features, which influence the potential performance of group strategies, are the validity of the group members' recognition and knowledge. Forty-three 3-member groups performed an inference task in which they had to infer which of two German companies had the higher market capitalization. Results based on the choice data support the hypothesis that groups adaptively apply the strategy that leads to the highest theoretically achievable performance. Time constraints had no effect on strategy use but on the proportions of different types of arguments discussed. Possible mechanisms that may underlie the adaptive use of recognition in group decision making are discussed.

---

[1] This chapter is in press in *Cognitive Science* under the same title.

Introduction

How do groups make decisions? And which factors influence their decision processes? Interestingly, the social psychological literature on group decision making is addressing very similar questions as the cognitive psychological literature that studies individual decision making, yet these two bodies of literature are as of now largely unconnected. Here, we show that it could be fruitful to combine the two to investigate an important aspect of group decision making: Do groups select decision strategies adaptively, given their composition regarding some task-relevant features, and how do they do so? The framework of ecological rationality has been applied successfully to study these questions in individuals, and here we show how it can be combined with classic models of group decision making to gain important insights for both research areas.

Teamwork has become a common form of organizational collaboration (Meyer, Shemla, & Schermuly, 2011; Salas, Cooke, & Rosen, 2008), especially to handle complex problems. Among the major advantages of groups are (1) that they possess the potential to pool more information and combine multiple perspectives (Larson, Foster-Fishman, & Keys, 1994; Stasser, 1992), (2) that work can be divided among several group members, and (3) that the work in groups is important for the satisfaction and well-being of people (Scholl, 2005). (4) Additionally, groups have a good potential to adapt to a (dynamic) environment (cf. Burke, Stagl, Salas, Pierce, & Kendall, 2006; Randall, Resick, & DeChurch, 2011). Research suggests, for example, that context variables and task demands affect strategy selection in groups. On the other hand, it has also shown that groups are far from being perfect in adapting their strategies. For instances, studies on the truth-wins principle suggest that groups often overlook a correct solution to a problem if the member who favors it cannot demonstrate its correctness (Laughlin & Ellis, 1986; Laughlin, VanderStoep, & Hollingshead, 1991). Moreover, research on social loafing (Karau & Williams, 1993) and hidden profiles (Stasser & Titus, 2003) showed that team work and decision processes in teams can also be dysfunctional (for a review on group performance, see Kerr & Tindale, 2004). For example people may be less motivated when working in a group than when working alone (Latané, Williams, & Harkins, 1979) or suffer from coordination losses (Steiner, 1972).

As the performance of teams can influence organizational success (Peterson, Owens, Tetlock, Fan, & Martorana, 1998), it is crucial to understand how groups make decisions. In particular, it is important to understand whether and how groups adapt their decision making process to both the composition of their group members and the structure of the task environment. Not all strategies are viable for all groups. Some strategies may need more expert knowledge than other, for instance. At the same time, not all strategies are viable for every particular task or situation. For example, there is ample evidence that the success of different decision strategies for combining individual inferences is conditional on certain task characteristics (Davis, 1992) and / or group composition features (Einhorn, Hogarth, & Klempner, 1977).

We propose the framework of ecological rationality to study the adaptivity of group decision making in detail. This framework assumes that people posses a repertoire of specialized decision strategies with which they can solve specific tasks in specific environments, which was coined the "adaptive toolbox" (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Goldstein, 1996; Gigerenzer & Selten, 2001; Gigerenzer, Todd, & the ABC Research Group, 1999). Importantly, none of the available strategies is an all-purpose tool that can be successfully applied to every situation. Rather, the success of a strategy is anchored both in the structure of the task environment and how much a strategy fits in this regard, and in the cognitive capabilities of the mind, both of which also determine which strategy is selected (e.g., Simon, 1956).

Groups can be conceptualized as information processing systems where cognition is distributed across individuals (Hinsz, Tindale, & Vollrath, 1997). We therefore propose that the notion of the adaptive toolbox can be transferred to groups as well. On the group level, the success of a strategy is still anchored in the structure of the task environment, as it is for individuals. But instead of the cognitive capabilities of one human mind, what matters on the group level are the cognitive capabilities of several human minds and their composition. In a different set of studies we have shown that groups adapt their decision strategies to the structure of the task environment (Kämmer, Gaissmaier, & Czienskowski, 2012). Here, we investigate whether and how groups adaptively select a particular decision strategy in relation to their composition with respect to a task-relevant feature. In doing so, we focus on a very simple decision strategy that has been studied extensively in individuals: the *recognition heuristic* (Goldstein & Gigerenzer, 2002).

Recognition heuristic: If one of two alternatives is recognized and the other is not, then infer that the recognized alternative has the higher value with respect to the criterion.

To clarify the terms used in this paper (see also, e.g., Reimer & Katsikopoulos, 2004), the recognition heuristic is only applicable in situations in which individuals recognize just one out of two objects. In these cases, the recognition heuristic predicts that the recognized object will have the higher value on some outside criterion. If both objects are recognized, people cannot rely on recognition to discriminate between the two, but they have to rely on further knowledge. The exact nature of this further knowledge and how it is applied is not further specified for the purpose here (see Goldstein & Gigerenzer, 2002). Thus, relying on knowledge in this context only means that both objects are recognized and are distinguished based on some other strategy than recognition. Finally, individuals who recognize neither object are assumed to guess.

Besides being well researched in individuals, we believe that this heuristic is highly suited to study adaptive decision making in groups, because it is (1) precisely specified, allowing to assess its success for each group and (2) has been previously shown to play an important role also in the context of group decision making (Reimer & Katsikopoulos, 2004).

We expect that for some groups it is more worthwhile to follow the recognition heuristic than for others. This allows us to find out whether groups relied on recognition in an adaptive manner. That is, do those groups for whom relying on recognition would be particularly successful rely more on it than groups for which other knowledge is better? Which variables influence whether a group relies heavily on recognition? And how can groups tell whether they should rely on recognition?

We will now first introduce research on the recognition heuristic in individuals before describing how it applies to the group context. Then, we will spell out in more detail what our specific questions and predictions are regarding adaptive decision making in groups.

The Recognition Heuristic in Individuals

The recognition heuristic is a prominent decision strategy in choice tasks for at least two reasons: First, recognition has a natural retrieval advantage over knowledge, which has to be searched for in memory with cognitive effort (Pachur & Hertwig, 2006). The recognition heuristic exploits the capacity for recognition, which is fundamental to the human mind (Standing, 1973). Due to the retrieval advantage, the recognition heuristic is often regarded as a default decision strategy (Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010). The second reason is an ecological one. Failure to recognize a name can be informative and facilitate decision making. Environmental mediators such as the media make it more likely that people encounter objects in their daily lives that are well-known and score high on a variety of criteria than those that score low. Consequently, objects scoring high on a criterion are often also more likely to be recognized. It was shown, for instance, that people are more likely to recognize larger cities (Goldstein & Gigerenzer, 2002), more successful political parties (Gaissmaier & Marewski, 2011), better colleges (Hertwig & Todd, 2003), and companies with high market capitalization (Marewski & Schooler, 2011). In domains with such interrelations and for people whose recognition rate highly correlates with the criterion (i.e., who have a high recognition validity, see Table 1), it is ecologically rational to rely on the recognition heuristic (Gigerenzer & Goldstein, 2011).

Over the last 10 years of research on the recognition heuristic, the conditions under which it is rational to rely on the heuristic have been specified and evidence has been provided that people adopt this heuristic in environments in which it yields accurate inferences (for a review, see Gigerenzer & Goldstein, 2011; for an overview of current studies, see special issues edited by Marewski, Pohl, & Vitouch, 2010, 2011a, 2011b; for a critical position, see Hilbig, 2010, and Pohl, 2011). People appear to select their strategies in an adaptive way by relying on recognition information when recognition is systematically related to a criterion but discounting it when it is not related (e.g., Pachur & Hertwig, 2006; Pohl 2006). For example, there is a positive correlation of $r = .64$ between the recognition validity and the proportion of judgments consistent with the recognition heuristic across 11 studies (Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011).

Recognition-Based Decision Models in Groups

To understand the decision process of groups better, it is important to clarify how groups are influenced by their members' recognition or lack thereof when forming joint judgments. Imagine a situation in which a number of people are asked to jointly infer which of two objects scores higher on a criterion. Some group members recognize only one of the two objects and form their inference on the basis of the recognition heuristic. Other members will only guess, while others again are more knowledgeable and they base their inference on additional information, which they retrieve from memory.

How do groups integrate their members' individual inferences in the described situation? The recognition heuristic applies to situations in which correct answers exist but where exact criterion knowledge is not available. Thus, an individual cannot demonstrate the correctness of a solution without additional sources, why this kind of task may be best described as having many features of a judgmental task (Davis, 1992; Laughlin, 1980), in which some form of a majority rule will best model the decisions of a group (Hastie & Kameda, 2005; Laughlin & Ellis, 1986). The most common majority rule is the simple majority rule, which infers that each group member has the same influence on a group decision and that the group chooses the option that receives the most votes. In addition to this rule, Reimer and Katsikopoulos (2004) proposed two oligarchic majority models that assign group members different weights in the voting process dependent on their individual decision strategy. The recognition-based model assumes that members using the recognition heuristic are more influential, while the knowledge-based model assumes that members using knowledge are (for further details, see Reimer & Katsikopoulos, 2004, pp. 1011). More specifically, (1) the recognition-based model (RBM) assumes that the simple majority of those group members who can use the recognition heuristic (i.e., those who recognize one but not the other object and choose the recognized one) determines the group choice, while all others are ignored, including those that have to rely on knowledge. (2) Conversely, according to the knowledge-based model (KBM), the majority of members who can use knowledge (i.e., those who recognize both objects) determines the group choice, while the opinions of all other members are ignored. Imagine, for example, a three-member group has to infer whether company A or company B has a higher market capitalization. Two members may have heard of A before, but do not recognize B, and thus infer (applying the recognition heuristic) that A has a higher market capitalization. The third member, on the contrary, recognizes both names and infers that B has the higher market capitalization based on some knowledge cues. According to RBM, the group choice would be A, while according to KBM, the group choice would be B. Note that both models are noncompensatory and predict that just one individual (in this case the knowledge-using member) can overturn a majority.

Reimer and Katsikopoulos (2004) empirically tested the models with three-member groups having to infer which of two American cities had the larger population. One of the major results was that, overall, RBM predicted the observed group choices better than KBM. This finding is surprising and counterintuitive as it shows that members lacking knowledge can dominate group decisions and trump a majority of members who recognize both objects and are, thus, more knowledgeable. This behavior—adhering more often to RBM than to KBM—was functional, though, because it increased groups' overall performance.

Despite this main finding of recognition dominating group decisions most of the time, in a considerable number of cases, recognition did not trump the vote of a majority of group members who could use their knowledge. A closer look at their results (see their Figure 3, p. 1019) reveals that there was a similarly large variance in strategy adherence among groups as has elsewhere been found in individuals (Pachur & Hertwig, 2006). Why groups relied on different strategies and whether they behaved adaptively in doing so, however, remained unexplored. We therefore aimed to study differences between groups who predominantly use RBM and those who predominantly use KBM and to explore strategy selection by testing the idea that groups select their strategies in an adaptive way. Extending this adaptiveness hypothesis to the group level, we assumed that groups are more likely to use RBM than KBM when RBM yields more accurate inferences, and vice versa. We next introduce the idea that groups, like individuals, adaptively select a strategy from their toolbox; we examine this hypothesis in more detail and discuss how time constraints may affect the adaptive selection of strategies.

The Adaptive Selection of Strategies in Group Decision Making

Given the adaptive use of the recognition heuristic in individuals, we aimed to test if groups are also able to select their strategies according to the theoretical accuracy of RBM and KBM. The theoretical accuracy of RBM and KBM informs about how often relying on the person using the recognition heuristic or using knowledge, respectively, will lead to correct inferences. As individual decision makers use the recognition heuristic in particular when recognition is a good predictor (termed item and environment adaptivity by Pachur, 2011), we expected similar findings for groups.

The quality of a group strategy (i.e., theoretical accuracy) likely depends on the average quality of individual strategies (here: recognition or knowledge validity) and / or of choices when decisions that were previously made on an individual level are integrated on the group level (e.g., Hastie, 1986; Laughlin, 1999). For example, groups are well advised to follow the opinions of their most knowledgeable members when these members are more often correct than their less knowledgeable members. We therefore expect that groups with a large number of members with high knowledge validity (or a high average knowledge validity per group) would rely predominantly on KBM. The opposite is expected for groups with a high average recognition validity. This latter assumption is supported by the study of Reimer and Katsikopoulos (2004) where the great adherence with RBM was accompanied by a high recognition validity and a low validity of group members' knowledge in every single group (on average .81 vs. .58, respectively). The reason for this uneven distribution was a methodological one: To be able to test for less-is-more effects, Reimer and Katsikopoulos (2004) had selected a subsample of American cities such that the validity of the recognition heuristic was considerably higher than the validity of group members' knowledge. Their study was the first showing that the less-is-more effect also exists on the group level, namely, that limited knowledge can result in a better inferential accuracy than more complete knowledge (Gigerenzer et al., 1999; Hertwig & Todd, 2003).

In sum, the major goals of the current study were to find out if groups apply KBM and RBM in an adaptive way and whether the use of the two different strategies is accompanied by a certain group composition regarding the knowledge and recognition validity. At the same time, we explored *how* groups might actually select strategies adaptively. In general, the influence of group members highly depends on the timing of contributions (Anderson & Kilduff, 2009) and confidence levels (Stasser & Davis, 1981). Here, it would be most adaptive if members with high recognition or knowledge validities had the strongest impact on the group decisions, respectively. Yet recognition and knowledge validities are most probably not directly accessible by individuals, so that they have to be inferred based on cues (Pachur, 2011). Potential cues include how quickly objects are recognized, how much other knowledge can be retrieved from memory or how trustworthy the source of recognition is evaluated.

How Is Strategy Use Reflected in Discussion Behavior?

An advantage of studying simple heuristics in a group setting is that people are urged to verbalize their strategies and reasons, making them easily observable. Although we cannot expect that verbalized reasoning exactly mirrors the strategies actually used (as those might not be accessible to people, e.g., Nisbett & Wilson, 1977), the process may still reveal participants' subjective, important reasons and provide a source for new insights for researchers (see Keller, Gummerum, Canz, Gigerenzer, & Takezawa, in press, for a similar claim). This may help to address the question of whether people really *use* the recognition heuristic (Hilbig, 2010), that is, whether they are relying on their recognition when the recognition heuristic models their choices best. Translated to the group level, the question is whether groups that predominantly use RBM also mention the recognition cue—constituting a behavioral correlate of RBM—more often during discussion than groups that predominantly use KBM (Reimer, Hertwig, & Sipek, in press). Thus, we measured the frequency with which the recognition cue was mentioned during discussion and also whether the recognition cue was perceived as a particularly valid argument compared to other pieces of information to test the hypotheses that the recognition cue is more frequently used by RBM than by KBM groups.

Another behavioral indicator of why groups predominantly adopt proposals by particular members may be derived by analyzing who speaks first when facing a new task. Previous studies have revealed that members speaking very early in a group discussion are more likely to exert influence than members speaking later (Abele, Stasser, & Groebe, 2012; Anderson & Kilduff, 2009; Shaw, 1961; Stasser, 2012). Therefore, we coded whether the member that could rely on the recognition heuristic or the member that has to rely on further knowledge spoke first, and hypothesized that in RBM groups the member using the recognition heuristic would more often contribute very early, while in KBM groups this should be the knowledge-using member.

Do Time Constraints Affect Strategy Use and Discussion Behavior?

Previous research on individual decision making has shown that time constraints have an impact on the extent to which different strategies are adaptive and on which strategies are used (e.g., Christensen-Szalanski, 1978, 1980; Marewski & Schooler, 2011; Pachur & Hertwig, 2006; Payne, Bettman, & Johnson, 1988, 1993; Rieskamp & Hoffrage, 2008; Svenson, Edland, & Slovic, 1990; Zakay, 1985). From group research and the attentional focus model (Karau & Kelly, 1992) we know that time constraints urge groups to focus on fewer and on more valid cues (Kelly & Loving, 2004). Because of its retrieval primacy, the recognition heuristic should become such a particularly valid cue. We thus included two time conditions (with and without time constraints) to explore the following assumptions. Time constraints should foster the use of RBM as RBM involves only one type of argument, the recognition heuristic, and should thus be a less time-consuming strategy than KBM, as KBM requires an exchange of arguments for each alternative. On the behavioral level, we expected to find an increased use of the recognition cue by RBM groups when they are under time constraints than when they are not as a sign of it being a particularly diagnostic cue for them.

Overview of Research Questions

In sum, we aimed to test whether groups select their strategies in an adaptive way, In addition to differentiating between groups that predominantly used RBM and groups that predominantly used KBM, we aimed to find out whether their choices led to the highest achievable accuracy. Moreover, we investigated whether time constraints affected the use of these strategies as opposed to no time constraints. Finally, to identify potential mechanisms behind adaptive strategy selection we assessed group composition features and behavioral correlates of the two decision strategies, which has rarely been done in previous research but is often asked for in the literature (Baumeister, Vohs, & Funder, 2007; Scholl, 2007).

Method

Participants and Design

One hundred thirty-two students (86 female, 46 male) from the Humboldt University Berlin, Germany, participated in the study. Their age ranged from 20 to 57 years with a mean age of 26.11 years (*SD* = 5.7). The three-member groups (*n* = 44) were randomly assigned to one of the two time conditions (with vs. without time constraints). Participants received performance-contingent payment for their participation and group performance in the economic comparison task [with an average compensation of about €17 ($26)]. One group had to be excluded from the analyses because of incomplete recording of individual answers. Thus, the final sample consisted of 43 groups (23 groups in the condition with time constraints and 20 groups in the condition without time constraints).

Experimental Task

In the economic comparison task, two German company names were presented and the one with the higher market capitalization had to be selected (task adapted from Marewski & Schooler, 2011, Experiment 5). The market capitalization reflects the current total value of a company and equates to the price a purchaser would have to pay for a full takeover (Glossary of Deutsche Börse Group, n.d.). Similar to the population comparison task used by Reimer and Katsikopoulos (2004), this is a magnitude inference task and rather a judgmental task (see introduction).

A sample of 100 company names, ranging from hardly to widely recognized, was selected from the 130 companies that were listed on the three major German stock exchanges in 2008 (DAX, MDAX, SDAX; see Table C.1 in Appendix C for the complete list).[2] On the basis of the individual recognition-test results, the 100 companies were randomly paired with each other once for each group separately for the paired-comparison task. The only restriction for the pairing was that the chance of obtaining pairs in which both RBM *and* KBM were applicable and differed in their predictions for the group choices (critical pairs) was maximized.[3] The resulting 50 pairs did not contain any company name twice. The two company names appeared simultaneously and in random position on the screen. Participants were asked to infer which company had the higher market capitalization. The correctness of answers was evaluated according to the market capitalization of the companies as of the month preceding the study. The term *market capitalization* was explained to the participants as defined above (see Appendix C for screenshots of the experimental paradigm).

Procedure

Upon arrival, participants were randomly assigned to one of the time conditions and to a three-member group. The experiment consisted of three parts, which lasted altogether approximately 1.5 hr. All tasks were administered through a PC. In the first part, participants individually completed a recognition test in which they were asked to indicate which of the 100 company names they recognized. The names were presented on a computer screen in random order and answers had to be indicated by pressing one of two keys (*recognition* or *no recognition*). Then, the paired-comparison task was administered individually. It contained 50 comparison pairs, being the same for the three participants that constituted a group in the second part, but not across groups.

For the second—the group—part, participants were asked to sit around a table facing a computer screen and a camera. Group members were instructed to come to joint decisions for the same 50 pairs of the individual paired-comparison task. They were asked to take turns typing the joint decision in a clockwise fashion so that no group member would play the role of a moderator or leader. Moreover, they were told that it was not possible to correct the decision after it was made and that they would not

[2] For the purpose of eliminating the completely unknown companies, we conducted a pilot study (that contained only a recognition test) with 40 lay people (24 females, mean age 23.5 years, $SD = 2.5$).
[3] This was achieved by basing the pairing on the recognition answers of each member and maximizing the numbers of pairs in which only two members recognized each company.

be given any feedback about the correctness of their answers until the end of the experiment. Finally, to increase their motivation, participants were informed that they would receive exclusively performance-contingent payment, namely, that each correct group decision would earn €0.50 ($0.78) for each person and that each member of a group would receive the same amount of money at the end of the experiment.

In the condition with time constraints, we restricted the discussion time of each comparison pair to 30s in total. The time a group had available to reach a decision was visualized by a countdown on the screen above the pair. The time between having answered one pair and releasing the next one (by pressing the space bar) was not restricted. Groups in the condition without time constraints were told that they could take the time they needed for their discussion.

In the third part, participants were asked to individually answer demographic questions and a manipulation check item: “I felt time pressure during the discussion” (answered on a scale from 1 *I totally disagree* to 5 *I totally agree*). Answers indicated that the respective time condition was successfully manipulated, $M_{\text{without time constraints}} = 1.30$, $SD = 0.58$ vs. $M_{\text{with time constraints}} = 3.56$, $SD = 1.18$; $t(117) = -12.78$, $p < .001$, $d = 2.44$. Last, they completed the *argument recall task,* which asked, “Which important arguments stated during discussion that spoke for or against a high market capitalization of a company do you recall? Please write down the four most important arguments speaking for a high market capitalization and the four most important ones that speak against it. Please rank them according to their importance (high scores indicate great importance).” Note that this recall task referred only to cues used as arguments and (in retrospect) to the complete discussion. Upon completion of the individual questionnaires, participants were debriefed, paid, thanked, and released.

Dependent Measures

All dependent variables are shortly explained in Table 1. From the answers to the two individual tasks (the recognition task and the paired comparisons), the dependent variables recognition and knowledge validity were calculated. To evaluate how well each decision model explained the group inferences, their predictive accuracies were calculated, that is, the proportion of trials in which the predictions of a model and the actual group choice converged. Moreover, we computed their theoretical accuracies, that is, the proportion of trials in which the models would choose the company with the higher market capitalization; and the achieved accuracy or performance of groups, that is, how often a group chose the company with the higher market capitalization.

Table 1

Dependent measures.

Measure	Explanation
Recognition validity α	$\alpha = R / (R + W)$, "where R is the number of correct (right) inferences the recognition heuristic would achieve, computed across all pairs in which one object is recognized and the other is not, and W is the number of incorrect (wrong) inferences under the same circumstances" (Goldstein & Gigerenzer, 2002, p.78).
Knowledge validity β	Proportion of correct answers when both objects are recognized
Predictive accuracy of RBM and KBM	Percentage of correctly predicted group choices a model makes for pairs in which it is applicable
Theoretical accuracy of RBM and KBM	Relative frequency with which a model yields correct choices
Achieved accuracy / performance	Observed number of correct choices per group

To explore how groups implemented their decision strategy, we analyzed the videotaped discussions and the answers to the argument recall task. The frequency of cues mentioned during discussion was determined on the basis of the Discussion Coding System (DCS; Schermuly & Scholl, 2011, 2012).[4] The general procedure of the DCS is to divide the interaction process into acts, which can be a sentence or several

[4] The DCS was applied by two trained coders. Six video-taped discussions of 30 min each were coded by both coders to determine the intercoder agreement [being κ = 0.82 for (1) *the recognition cue* and κ = 0.83 for (2) *knowledge cues*], which was satisfactory and thus justified the decision to have the remaining 37 discussions coded by only one coder each. Due to incomplete or partly damaged videotapes, approximately 3% of the observation data (66 of 2,150 comparison pairs) are missing.

sentences of the same topic, and then to code the functional and interpersonal meaning and content of each act. The content of an act was coded into two main categories:
(1) An act was coded as a *recognition cue* when a person used the recognition heuristic as an argument or simply let the other members know that she/he recognized one company name but not the other. It was also coded when a person used the recognition cue of another member or the whole group as an argument. Example statements are, “If you do not even recognize company A, it cannot be big,” and “Let’s take company A, since we all recognize it.” (2) Acts were coded as *knowledge cues* when a group member provided cues about a company other than recognition. Again it did not matter whether a person simply stated his/her cue knowledge (e.g., “I know that company B produces drugs”) or used a cue as an argument for or against a high market capitalization (e.g., “Company A is a bank, and banks have money”), although the latter was much more often the case (the same holds for category 1). Note that category 2 is wider than the recognition-cue category as it contains different kinds of cues. To evaluate the answers to the *argument recall task*, all (pro and con) arguments that included “(no) recognition” or “(no) renown” were coded as *recognition cues* together with their ranking (0 *not mentioned*, 1 *lowest rank*, 4 *highest rank*).

Results

Individual Condition

In their individual condition, participants recognized on average 46 out of 100 companies (SD = 11.99; range 22–84) and were correct in 60.9% of the 50 inferences (SD = 8.4, min = 40%, max = 82%). Confronted with pairs where they recognized both companies, they made 66.2% correct inferences (SD = 15.5), that is, their average knowledge validity was .66. With pairs where they recognized neither company they had an average guessing accuracy of 49.4% (SD = 13.5). Finally, with pairs where they recognized just one company, they made 71.5% correct inferences (SD = 14.7). Their average recognition validity was .69. Thus, different from the study by Reimer and Katsikopoulos (2004), the average recognition and knowledge validities were very similar to each other in this study. On average, participants individually adhered to the recognition heuristic in 83.0% of situations (SD = 12.8) in which it was applicable.

Group Condition

All groups together accomplished 43 × 50 = 2,150 inferences. On average, groups answered 66.4% (SD = 9.4, min = 44%, max = 84%) of all choices correctly. We first tested whether the two time conditions had an impact on the predictive accuracies of RBM and KBM. There were no differences regarding the predictive accuracies or RBM and KBM between the time conditions, respectively, neither for RBM: $t(41) = 0.003$, $p = .997$, $d = 0.001$, nor for KBM: $t(41) = 0.818$, $p = .42$, $d = 0.25$.[5] Therefore, the subsequent analyses concerning strategy use are based on the joint data set from both time conditions, consisting of 43 groups or 2,150 comparison pairs.

We next analyzed how often each strategy was applicable, and how often the groups' choices were in accordance with each strategy. The simple majority rule (SM) was applicable to all 2,150 cases and 74.6% of its predictions matched the group choices. Note that, as Reimer & Katsikopoulos (2004), we will focus on RBM and KBM rather than SM subsequently, as they and their correlates will be more informative regarding adaptive strategy selection. RBM was applicable to 1,243 cases (57.8%) and 74.7% of its predictions agreed with the group choices. KBM could be applied to 1,015 cases (47.1%), and 80.6% of its predictions agreed with the group choices.[6] The higher overall predictive accuracy of KBM was functional because the overall theoretical accuracy of KBM was higher (72.9%) than that of RBM (70.8%). As a first hint of adaptive strategy selection, the overall predictive accuracy of KBM (RBM) on the group level was positively correlated with the theoretical accuracy of KBM (RBM), $r_{KBM}(41) = .23$, $p = .13$, $n = 43$; $r_{RBM}(41) = .39$, $p = .01$, $n = 43$. That is, groups for whom a strategy yielded more accurate inferences also tended to adhere to that strategy more often, which held particularly true for RBM.

Because the two subsamples in which RBM and KBM were applicable only overlapped partially, we also considered the subset of situations where both restricted

---

[5] None of the dependent variables was affected by the time condition, except the behavioral correlates (see below), as a multivariate ANOVA with the predictive and theoretical accuracies of RBM and KBM, respectively, and the achieved accuracy per group as dependent variables, and the time conditions as independent variable revealed, $F_{time}(5, 37) = 0.700$, $p = .63$, $\eta_p^2 = .09$.

[6] Both models fail to make predictions if there is no majority of one kind, but members using the same individual strategy contradict each other. In 14 of the 907 trials (1.5%) in which RBM made no prediction (2,150 – 2,143) RBM failed to make a prediction because there was a contradiction between two recognition heuristic users. In 225 or the 1,135 trials (19.8%) in which KBM made no prediction there was a contradiction between two knowledge users.

majority models were applicable (n = 811). Here, 81.0% of the predictions made by KBM and 74.4% of those made by RBM agreed with the group choices. Again, the theoretical accuracy of KBM was higher (73.4%) than that of RBM (71.9%; for correlations see below). An analysis of the 181 situations in which the two decision schemes made contrasting predictions (critical pairs) showed that KBM matched the group choices in 115 situations (64%), and RBM in 66 (36%). In sum, aggregate analyses suggested a higher predictive accuracy of KBM overall, together with a higher theoretical accuracy of KBM. If not stated otherwise, the following analyses were based on the subset of 811 situations in which both models could be applied.

Classification of groups as RBM or KBM groups

Figure 1 depicts the predictive accuracies of the two strategies for each individual group, ordered according to the predictive accuracy of RBM. It shows that the overall differences between the predictive accuracies of KBM and RBM per group are not negligible but range between 3.7 and 50 percentage points (M = 14.5, SD = 8.9). The large variance of the predictive accuracies of KBM and RBM suggests differences in the strategy use between groups. As with individuals, different groups obviously preferred different strategies and differed in the extent to which they adhered to one of the two strategies.

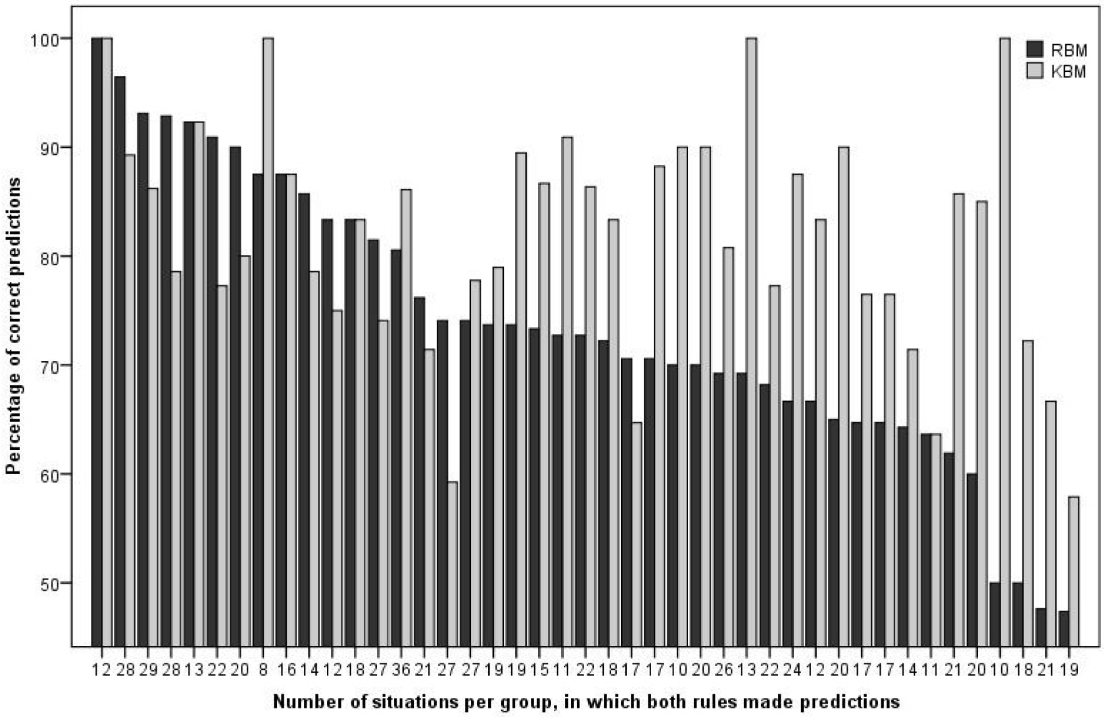


Figure 1. Percentage of correct predictions by the knowledge-based model (KBM) and the recognition-based model (RBM), for each group (bars indicate the percentage of situations in which KBM and RBM made correct predictions for the group choices). Groups are sorted in a descending order according to the predictive accuracy of RBM. Along the *x*-axis, the total number of situations per group in which both models, KBM and RBM, made predictions, are displayed (overall $n = 811$).

To test the adaptiveness hypothesis our next step was to classify groups. Several studies on heuristic decision making have argued that systematic individual differences in decision strategies are rather the rule than the exception, and thus, only analyses that take the individual as the unit of analysis allow for a proper assessment of the underlying processes of cognitive strategies (e.g., Gigerenzer & Goldstein, 2011; Marewski, Schooler, & Gigerenzer, 2010; Pachur, Bröder, & Marewski, 2008). Thus, in addition to comparing mean values, we conducted analyses on the level of individual groups by classifying each group as an RBM group when more of its choices were matched by the predictions of RBM than by KBM (i.e., if the predictive accuracy of RBM was higher than that of KBM) and vice versa. As shown in Figure 1, the group choices of 27 groups could be best explained by KBM with a mean predictive accuracy of 83.7%, as opposed to a mean predictive accuracy of 66.9% of RBM, paired $t(26) =$

9.187, $p < .001$, $dz = 1.77$. We classified these groups as KBM groups. Conversely, the data of 11 groups could be best explained by RBM with a mean predictive accuracy of 85.0%, as opposed to a mean predictive accuracy of 75.9% of KBM, paired $t(10) = -8.517$, $p < .001$, $dz = 2.56$. We classified those groups as RBM groups. The data of five groups could be equally well described by both models (with a mean predictive accuracy of 85.4%). These five groups were excluded from further analyses, so that the following results are based on the 38 classified groups.[7] Table 2 summarizes all key differences between RBM and KBM groups reported so far and subsequently.

Table 2

Summary of key differences between RBM and KBM groups, summarized across time conditions (means, SD in parentheses).

	RBM groups	KBM groups
N	11	27
Predictive accuracy of RBM[1]	85.0% (8.6)	66.9% (9.5)
Predictive accuracy of KBM[1]	75.9% (8.6)	83.7% (9.8)
Theoretical accuracy of RBM[1]	78.8% (8.1)	67.3% (7.4)
Theoretical accuracy of KBM[1]	71.8% (9.9)	72.0% (11.6)
Achieved group accuracy[1]	80.3% (9.7)	73.7% (12.3)
Individual recognition validity	.74 (.13)	.66 (.13)
Individual knowledge validity	.66 (.12)	.67 (.15)
Individual speed of recognizing a company (in msec)	1,551 (408)	1,803 (623)
Individual speed of *not* recognizing a company (in msec)	1,556 (376)	1,738 (668)
Accordance with recognition heuristic when it led to a		
- correct choice[1]	91.1% (7.3)	80.1% (15.5)
- incorrect choice1	59.5% (19.8)	38.9% (15.1)
Relative frequency of speaking first of		
- member using the recognition heuristic[1]	53.4% (17.0)	45.1% (17.3)
- member using knowledge[1]	34.7% (17.4)	49.2% (17.8)

[1] Based on 811 pairs where strategies made different predictions.

[7] We run the same classification procedure also including the predictive accuracy of the simple majority rule (SM) into the comparison. Ties between the predictive accuracies of RBM or KBM and SM respectively were resolved in advantage for RBM or KBM. This classification resulted in 10 RBM groups, 24 KBM groups and 6 SM groups. Using those 34 RBM and KBM groups rather than all 38 groups yields the same results as all subsequent analyses.

Was the strategy choice of groups adaptive?

To test the assumption that groups behaved adaptively, we analyzed the theoretical accuracies of the two models for KBM and RBM groups. We first conducted a second classification of all groups based on the theoretical accuracies of the two models. As with the predictive accuracies, the analysis compared the theoretical accuracies of the two models for each group and classified a group as a *t*-RBM group (*t*-KBM group) if the theoretical accuracy of RBM (KBM) was higher. Figure 2 depicts the theoretical accuracy of the two strategies for each group (and also the achieved accuracy, to which we refer in the next section). As expected, the theoretical accuracies differed between groups, which suggests that different strategies were adaptive for different groups: Of the 38 groups, 18 were classified as *t*-KBM groups (for whom the theoretical accuracy of KBM was on average 12.0 percentage points higher than that of RBM) and 14 as *t*-RBM groups (for whom the theoretical accuracy of RBM was on average 11.8 percentage points higher than that of KBM). For six groups the theoretical accuracies were identical.

If strategy selection were adaptive, the classification of groups based on predictive accuracy (i.e., their choices) should be similar to the classification based on the theoretical accuracy of the strategies. If we look at those 32 groups for which the theoretical accuracies allowed a classification, we find a convergent classification rate of 71.9% (23 out of 32) with the classification based on the predictive accuracies. The theoretical accuracy of KBM (RBM) was positively correlated with the predictive accuracy of KBM (RBM) on the group level, $r_{KBM}(36) = .39$, $p = .02$, $n = 38$; $r_{RBM}(36) = .48$, $p = .003$, $n = 38$.

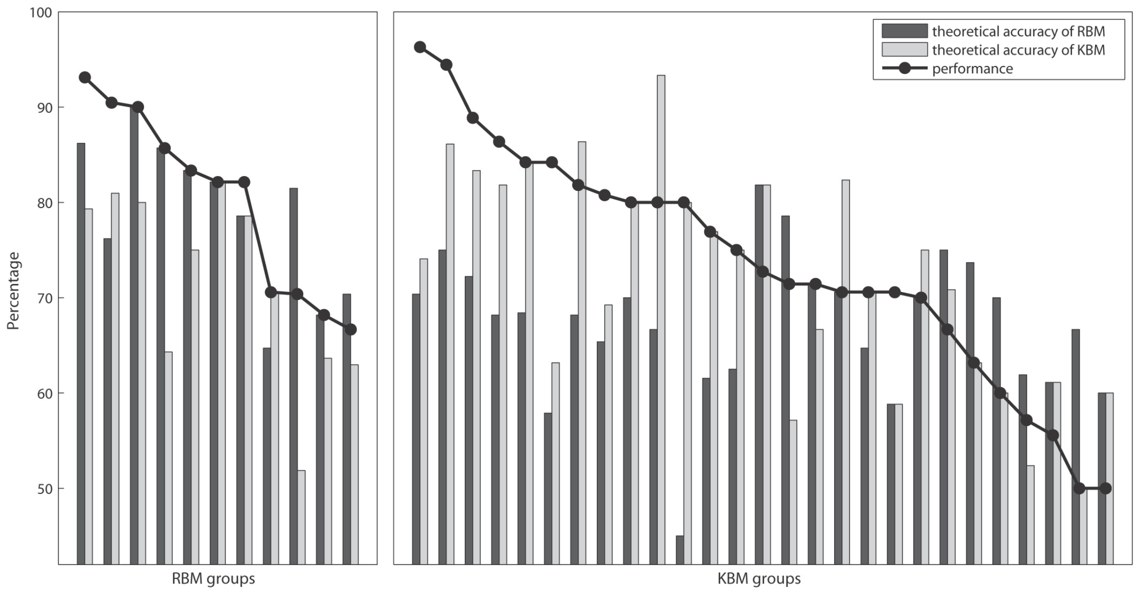


Figure 2. Achieved accuracy and theoretical accuracy of KBM and of RBM for each of the 38 categorized groups in the subset of situations in which both models were applicable (n = 811). The graph displays the values of RBM groups on the left and those of KBM groups on the right. Within these categories, groups are sorted according to their achieved accuracy in descending order. It can be seen that for RBM groups, the theoretical accuracy of RBM was higher and closer to the observed performance than the theoretical accuracy of KBM, and vice versa for KBM groups.

Adaptive behavior should be mirrored by a high performance level. To look into this aspect, we compared the achieved accuracy per group with the accuracy that would have been achievable if groups had consistently used one of the two decision strategies (i.e., the *theoretical accuracies*; see Figure 2). These analyses were also run for the subset of situations in which both models could be applied. On average, we observed a mean achieved accuracy of 80.3% for the 11 RBM groups. The mean theoretical accuracy of RBM for RBM groups was 78.8% and of KBM only 71.8%, paired $t(10) = -2.192$, $p = .05$, $dz = 0.37$. The 27 KBM groups made fewer correct choices in the subset than RBM groups, namely, 73.7%, on average.[8] The mean theoretical accuracy of KBM for KBM groups was 72.0% and of RBM only 67.3%, paired $t(26) = 1.942$, $p = .06$, $dz = 0.66$. Thus, as the classification of the groups on the

[8] One reason could be that KBM group members received more difficult pairs by accident. This was, however, not the case: The mean differences between the market capitalizations of the two companies of a pair were equal for KBM and RBM groups (M_{KBM} = 11,242,721,413, SD = 17,046,506,195 vs. M_{RBM} = 11,535,090,275, SD = 17,261,293,274), $t(1898) = -0.338$, $p = .74$, $d = -0.02$.

basis of the theoretical accuracies had already suggested, these results support the conclusion that groups chose their strategies in a majority of trials in an adaptive way by picking the strategy that yielded the higher accuracy.

Mechanisms behind adaptive strategy use

One determinant of the quality of group strategies (i.e., their theoretical accuracy) might be the quality of individual members' strategies and choices (i.e., the knowledge and recognition validity). In fact, the theoretical accuracy of KBM was highly correlated with the average knowledge validity ($r(36) = .65$, $p < .001$, $n = 38$) of the members of a group; similarly, the theoretical accuracy of RBM was highly correlated with the average recognition validity ($r(36) = .74$, $p < .001$, $n = 38$). Interestingly, in RBM groups, the members' average recognition validity ($M = .74$, $SD = .13$) was considerably higher than their knowledge validity ($M = .66$, $SD = .12$), as in the study by Reimer and Katsikopoulos (2004). Conversely, in KBM groups, the average recognition ($M = .66$, $SD = .13$) and knowledge ($M = .67$, $SD = .15$) validity were very similar to each other.

One possible cue to one's knowledge and recognition validities are one's recognition times (Pachur, 2011). An ANOVA with repeated measurements with the individual recognition judgment (recognized vs. not recognized) as the within-subjects factor and the type of group as the between-subjects factor revealed that RBM group members had indeed faster recognition times than KBM group members (recognition judgments: $M_{\text{RBM}} = 1{,}551$ msec, $SD = 408$ vs. $M_{\text{KBM}} = 1{,}803$ msec, $SD = 623$; no-recognition judgments: $M_{\text{RBM}} = 1{,}556$ msec, $SD = 376$ vs. $M_{\text{KBM}} = 1{,}738$ msec, $SD = 668$; $F_{\text{group type}}(1, 112) = 3.59$, $p = .06$, $\eta_p^2 = .03$. Thus, RBM-group members may have inferred their higher recognition validity from their faster recognition times.

Adaptation on individual trials within groups

It is a simplifying assumption that groups use one strategy consistently, which is however necessary for the classification and to derive conclusions about differences between groups. Nevertheless, it is of course possible that groups also adapt their strategy to individual trials. An interesting measure of strategy adaptation on the trial level is to distinguish cases in which the recognition heuristic leads to a correct decision and those where it leads to an incorrect decision (the difference of these two proportions is termed discrimination index; Hilbig & Pohl, 2008). Members of RBM and KBM groups both *individually* accorded with the recognition heuristic to a similar degree when it led to a correct as when it led to an incorrect decision (RBM-group members: M_{correct} = 93.6%, *SD* = 11.9 vs. $M_{\text{incorrect}}$ = 88.4%, *SD* = 19.9; paired $t(29)$ = 1.445, p = .16, dz = 0.26; KBM-group members: M_{correct} = 91.0%, *SD* = 15.7 vs. $M_{\text{incorrect}}$ = 86.8%, *SD* = 22.9; paired $t(63)$ = 1.300, p = .20, dz = 0.16).

Groups, in contrast, showed a more selective accordance with the recognition heuristic and accorded with it more often when it led to a correct decision than when it led to an incorrect decision, and this was more strongly the case for KBM groups than for RBM groups (RBM groups: M_{correct} = 91.1%, *SD* = 7.3 vs. $M_{\text{incorrect}}$ = 59.5%, *SD* = 19.8; paired $t(10)$ = 5.115, $p < .001$, dz = 1.54; KBM groups: M_{correct} = 80.1%, *SD* = 15.5 vs. $M_{\text{incorrect}}$ = 38.9%, *SD* = 15.1; paired $t(26)$ = 8.659, $p < .001$, dz = 2.03). Thus, both RBM and KBM groups showed a more selective accordance with the recognition heuristic than individuals, indicating that groups incorporate information beyond recognition more strongly than individuals. Congruent with the classification of groups as RBM or KBM group, the selectivity of accordance was much less pronounced in RBM groups than in KBM groups: Only in RBM groups, the majority of choices (i.e., > 50%) less selectively accorded with the recognition heuristic both when it led to a correct as well as when it lead to an incorrect decision, which was not true for KBM groups.

Behavioral correlates of RBM and KBM

Did RBM groups differ in their group discussions from KBM groups? We first focused on the very first act of the discussion concerning each trial. Independent of the content of the first argument, we coded who spoke first. Did members who could rely on the recognition heuristic predominantly speak first in RBM groups, while members who could rely on knowledge did so in KBM groups? In fact, members who could rely on the recognition heuristic spoke first slightly more frequently in RBM groups (M = 53.4%, *SD* = 17.0) than in KBM groups (M = 45.1%, *SD* = 17.3), $F_{\text{group type}}(1, 34)$ = 1.806, p = .19, η_p^2 = .05., while those members who could rely on knowledge spoke first more often in KBM groups (M = 49.2%, *SD* = 17.8) than in RBM groups (M =

34.7%, SD = 17.4), $F_{\text{group type}}$ (1, 34) = 6.559, p = .02, η_p^2 = .16. No differences between the time conditions were revealed.[9] Speaking first thus nicely mirrored the group strategy on the behavioral level.

We then analyzed the arguments exchanged during the decision process. Did RBM groups mention the recognition cue more often than KBM groups in general? And did RBM groups under time constraints mention the recognition cue more often than RBM groups without time constraints? First, we computed the average joint number of all recognition and knowledge cues that were exchanged per trial (within the subsample of 811 trials). We then entered this number into an ANOVA with the two time conditions and two group types as independent variables. No difference was revealed between the group types, $F_{\text{group type}}$ (1, 34) = 1.00, p = .33, η_p^2 = .03, but—quite naturally—there was a difference between the time conditions: Groups with time constraints discussed fewer cues (M = 2.39, SD = 0.77) than groups without time constraints (M = 3.97, SD = 1.79), F_{time}(1, 34) = 7.12, p = .01, η_p^2 = .17. Next, we computed the relative frequency of the recognition cue with regard to all cues and took this as the dependent variable in a second ANOVA. It revealed a main effect of time, namely, that the relative frequency of the recognition cue in groups with time constraints was higher (M = 14.3%, SD = 8.1) than in groups without time constraints (M = 9.4%, SD = 5.6), F_{time}(1, 34) = 10.56, p = .003, η_p^2 = .24; in other words, groups exchanged a greater proportion of knowledge cues when they had time. More interestingly, it revealed that, RBM groups under time constraints mentioned the recognition cue much more often (M = 20.3%, SD = 7.4) than RBM groups without time constraints (M = 6.36% , SD = 3.24) and than KBM groups in both time conditions ($M_{\text{with time constraints}}$ = 10.7%, SD = 6.4; $M_{\text{without time constraints}}$ = 10.4%, SD = 5.8), $F_{\text{Time x Group type}}$(1, 34) = 9.51, p = .004, η_p^2 = .22.

The results of the *argument recall task* show a similar picture. Recall that every participant was asked to name arguments pro *and* con, so that the recognition cue (recognition or no recognition) could be named zero times, once, or twice per person. For the following analyses, we counted the number of people who recalled the

[9] Mere differences in the base rate of members using the recognition heuristic (R members) or knowledge (K members) in the first trial between RBM and KBM groups cannot account for this result [R members: $F_{\text{group type}}$ (1, 34) = 1.384, p = .25, η_p^2 = .04; K members: $F_{\text{group type}}$ (1, 34) = 0.721, p = .40, η_p^2 = .02]. There were also no differences between time conditions [R members: F_{time} (1, 34) = 1.380, p = .25, η_p^2 = .04; K members: F_{time} (1, 34) = 0.258, p = .62, η_p^2 = .01].

recognition cue at least once and found that 25 out of 33 RBM-group members (75.8%) and 59 out of the 81 KBM-group members (72.8%) did so. Figure 3 depicts the percentage of times the recognition cue was not mentioned (0) and rated as low (1) to high (4) importance by people who recalled the recognition cue at least once. It shows that, in the condition without time constraints, the recognition cue was ranked more often as of high importance by RBM-group members and was more often not mentioned by KBM-group members. In the condition with time constraints, more RBM-group than KBM-group members rated the recognition cue as of moderately high (3) and high (4) importance, while the reverse was true for the low (1) and moderately low (2) importance ratings. To summarize, the observed frequency and importance ratings of the recognition cue indicate that the recognition cue was perceived as a valid cue and is a behavioral correlate of RBM, under time constraints in particular.

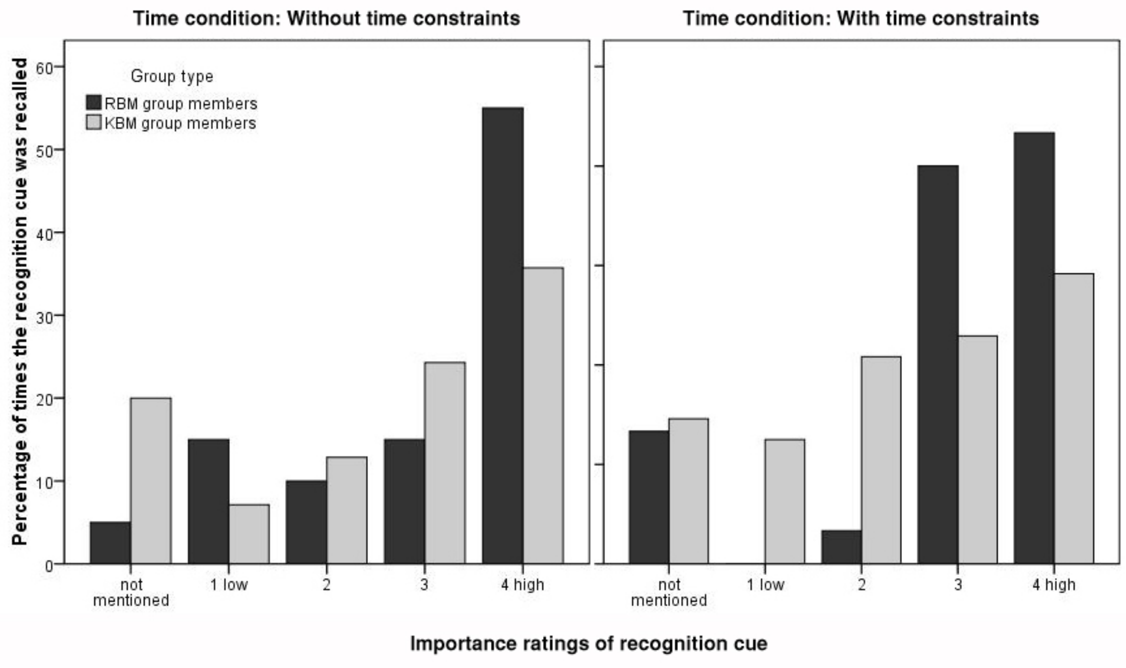


Figure 3. Percentage of times the recognition cue was recalled by participants who mentioned it at least once in the argument recall task (n = 84), on the left for the condition without time constraints (n = 45) and on the right for the condition with time constraints (n = 39).

Discussion

In many instances in everyday life as well as in organizations, decisions are made by groups (Salas et al., 2008). Here, we applied the framework of ecological rationality to investigate an important aspect of group decision making: Are groups able to select strategies adaptively? From this perspective, transferred from the individual to the group level, the success of a decision strategy is anchored both in the structure of the task environment and in the composition of the individual minds of the group. We focused on the second aspect, that is, on whether and how groups adapted their strategy to group composition.

Summary of Results

To this end, we studied two particular decision strategies: one restricted majority rule that assumes that only the votes of group members that can rely on knowledge determine the group decision (KBM); and another that assumes that only the votes of group members that can rely on the recognition heuristic are taken into account (RBM). Our main purpose was to test if groups select these two strategies in an adaptive way in a paired comparison task. Adaptivity was defined as the degree to which the strategy was chosen that also yielded the more accurate inferences. In fact, we found that the choices of most groups were best predicted by the strategy that also led to the highest theoretical accuracy. Therefore, groups, on average, achieved a performance level that was close to the best they could achieve theoretically (see Figure 2).

In addition to our general question on adaptive strategy selection, we tested the effect of time constraints on strategy use. Contrary to our expectations, we did not find that time constraints led to an increased use of RBM (no difference in predictive accuracy of RBM between time conditions). Our rationale behind this assumption was that RBM would require fewer arguments than KBM, rendering it a faster and more frugal strategy. No differences in the total number of arguments, however, were found between strategies. Consequently, on first sight, this result seems to contradict findings in individual decision strategy research showing that the use of noncompensatory decision rules increases with decreasing time available (e.g., Christensen-Szalanski, 1980; Pachur & Hertwig, 2006; Svenson et al., 1990; Zakay, 1985). On second sight, however, this result is plausible for two reasons. First, on the group level, one can regard both RBM and KBM as noncompensatory strategies in that a minority of group members can trump a majority. Second and more importantly, the indifference in predictive accuracies of RBM and KBM supports the adaptiveness findings, as KBM and RBM did not differ in their theoretical accuracies between the time conditions in the first place (see footnote 4). In other words, RBM would not have been more adaptive under time constraints. Time constraints in the group phase may not have influenced decision strategies anymore, because opinion formation had already taken place individually preceding the group discussion, and there were no time constraints in this individual phase of the experiment. It would be an interesting question for future research to test whether time pressure faced by individuals who make a decision prior to meeting in a group has an impact on the group strategy.

Interestingly enough, we found an effect of time constraints on discussion behavior, namely, that time constraints led to an increase in the relative frequency of the recognition cue being discussed as compared to knowledge cues by all groups and in particular by RBM groups. This enhancement effect seems to support the assumption that recognition is a very valid cue (Kelly & Loving, 2004) and that it plays a special role in decision making. Ratings of the freely recalled arguments again revealed the higher importance of the recognition cue for RBM groups, as more RBM-group members than KBM-group members mentioned and perceived the recognition cue as highly important. The number and persuasiveness of arguments may cause a group to shift to an alternative that has not been favored by a majority before (Hinsz & Davis, 1984). Despite these supportive differences in the relative frequencies, the absolute numbers reveal that mentions of the recognition cue constituted only a small proportion of all mentioned cues. A technical reason for this finding may be that the categories had different widths: While the recognition cue was counted only when it was mentioned as such, the knowledge-cues category comprised all cues containing information about the companies at hand, such as the company's sector or products. Another reason might be seen in the justification pressure caused by the group setting, which often leads people to use more information than they initially used in individual decisions (Huber & Seiser, 2001; Lerner & Tetlock, 1999).

As a final behavioral correlate, we assessed who spoke first and found that this was indeed an indicator of the group strategies used, with members classified as knowledge users speaking more often first in KBM groups and vice versa in RBM groups. In addition to regarding speaking first simply as a behavioral correlate, it may also be seen as a process measure and a reason why groups end up with using one strategy or another (cf. Stasser, 2012). Indeed it has been found elsewhere that the first answers provided have a strong impact on the final group answer (Anderson & Kilduff, 2009). More research is needed to tease apart whether contributing early is only an indicator or a driving force for selecting a certain decision (rule). In sum, we found evidence for the adaptive selection of two strategies in small groups, that time constraints had no impact on this selection but enhanced the usage of one of the most valid cues, the recognition cue.

Mechanisms Behind Adaptive Strategy Selection

How did group members decide whether to follow those members who could rely on the recognition-heuristic—in an adaptive way? These questions rephrase a question that is also central to the concept of the recognition heuristic as part of the adaptive toolbox for individual decision making. More generally, this fundamental problem of how people decide how to decide is known in the literature as the strategy selection problem, and a number of approaches to strategy selection have been proposed (e.g., Payne et al., 1988, 1993; Rieskamp & Otto, 2006; Scheibehenne, Rieskamp, & Wagenmakers, 2012; for an alternative account see Newell & Lee, 2011; for a comment see Cooper, 2000; for a debate see Glöckner, Betsch, & Schindler, 2010; Marewski, 2010). Note that we do not claim that people have direct access to the theoretical accuracies of different strategies and then consciously decide which one to follow (neither is this an assumption of any of the strategy selection accounts). Rather we hypothesize some plausible mechanisms that lead groups on a trial-by-trial basis to adopt a certain choice and that leads them—with regard to the accomplishment of the complete task—to appear to predominantly follow one rule.

In our case, one scenario could be that KBM serves as a default strategy, as “more knowledge” (here: recognizing both objects rather than just one) is typically assumed to be better. Something similar was observed in individuals who often start out with decision strategies that consider all pieces of information (see e.g., Bröder & Schiffer, 2006). This would mean that groups pool their available knowledge and follow the most knowledgeable member(s) by default, and only follow members who can rely on the recognition heuristic (and who thus only recognize one object) if they have strong reasons to do so. Such reasons could be that (1) the knowledge cues of knowledge-users may be in fact weaker than the recognition cue, or that (2) the knowledge cues do not suggest a direction for the decision (such as when people know that A is a bank and B is a car producer but do not know which sector scores higher on the criterion market capitalization), while the recognition heuristic always entails the direction. (3) A third possible reason is based on the assumption that the use of the recognition heuristic involves two distinct processes, namely the judgment whether an object is recognized or not and second the evaluation of whether recognition is a useful and reliable indicator given the task (see Volz et al., 2006, for evidence for the neural basis of these two processes). People may thus consider, for example, the source of recognition to evaluate the validity of their recognition (see Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2009; Pachur, 2011). In the group setting, they may then communicate to the others their evaluation result by either directly naming the source or indirectly through their confidence level or other paralinguistic signals.

An indicator for the plausibility of this default scenario may be seen in the large proportion of KBM groups in our study (27 to 11 RBM groups). The differences between the average group recognition and knowledge validities supports the idea that knowledge is only abandoned when the validity of recognition is substantially larger, as it amounted to only 1 percentage point for KBM groups on average but to 8 percentage points for RBM groups (in favor of the recognition validity). Note that this latter difference is still smaller than it was in the groups studied by Reimer and Katsikopoulos (2004) where the average difference between recognition and knowledge validities was as high as 23 percentage points, which could explain why the RBM played a stronger role there.

A second possible scenario may be outlined as follows: group members have access to the quality and adaptivity of their recognition and knowledge, which they communicate to the other members, for example, by their confidence or knowledge cues, or by the timing of putting forward their arguments (namely by speaking first; e.g., Anderson & Kilduff, 2009), and by this lead the group decision into one direction.

How could people access the quality of their recognition? One hypothesis concerns the role the recognition speed (i.e., fluency) might play. As Schooler and Hertwig (2005) argued, recognition speed can be taken as a proxy for the activation of a memory record: “the lower the activation, the more time it takes to retrieve a record” (p. 616). Activation of a memory record is a noisy process. The clearer the activation, the more systematic the image of the world around us should be, and thus the better people should be able to tell a truly recognized from a truly unrecognized object, rendering recognition a diagnostic cue (for the impact of additional knowledge in recognition cases on decision times see Hilbig & Pohl, 2009). If we now assume that people are sensitive to differences in recognition times and that fluency is often correlated with the criterion (Hertwig, Herzog, Schooler, & Reimer, 2008; Schooler & Hertwig, 2005), it seems plausible that people may (correctly) rely on fluency to infer their recognition accuracy (i.e., validity). So far, this mechanism has been proposed to explain the adaptive use of the recognition heuristic on the level of a single item (item adaptivity) as well as on the level of a particular environment (environment adaptivity) by individual decision makers (Pachur, 2011; Pachur & Hertwig, 2006). We propose that it might also play a role on the group level, since fluency is also important in social interactions, because short recognition times elicit high confidence (Zakay & Tuvia, 1998) or indicate certainty (Erdfelder, Küpper-Tetzel, & Mattern, 2011), and greater confidence makes people more influential in a freely interacting group (Thomas & McFadyen, 1995; Zarnoth & Szniezek, 1997). Indeed, we found that RBM-group members gave slightly faster judgments in the recognition task than KBM-group members, which may have rendered them more confident and thus influential, causing those groups to rely more often on RBM. Notwithstanding the plausibility of the aforementioned possible mechanisms underlying the adaptive use of RBM and KBM, they deserve more future research.

Limitations

This study is of course only a first step towards studying adaptive strategy selection by groups. Its biggest limitation is its focus on one particular strategy, the recognition heuristic, which is only applicable in a limited set of situations. On the other hand, the heuristic and the tasks to which it can be applied are very well defined, and the heuristic can be applied on the individual level and transferred to the level of groups. Despite its limited applicability, these aspects were ideal preconditions for testing adaptive strategy selection on the group level in relation to the composition of groups. Nevertheless, it would be important to expand this line of research to different models of group decision making and to different tasks.

Another limitation of our study was that we studied ad-hoc groups that mainly consisted of students who were unfamiliar with the task. While this procedure had methodological advantages, it limits the generalizability of the findings. It would thus be important to study real groups in the future, with tasks or contexts they have acquired some expertise in.

Implications

The study demonstrates that the framework of ecological rationality is fruitful to study adaptive strategy selection in groups. As Pachur and colleagues (2008) concluded, individual differences in cognitive strategies appear to be the rule rather than the exception (see also Gigerenzer & Goldstein, 2011). This is transferable to groups as well. By focusing on the group level and intergroup differences as an addition to aggregate analyses, the current study contributes to a more “ideographic” type of research as opposed to more “nomothetic” research, which focuses on identifying general laws of information processing. With the attempt to capture dependencies between decision strategies and group structures, the current study fits to the adaptive toolbox approach and extends it to the group level. It also demonstrates that group research can benefit from formal model testing as it allows to empirically test different models on a trial-by-trial basis and to derive quantitative evaluations of competing models.

In this way it links to (1) team composition research, which focuses on the impact of team composition regarding surface (i.e., overt demographic characteristics) as well as deep-level attributes (such as personality traits or general mental ability) on processes and team effectiveness, especially in an organizational context (e.g., Bell, 2007; Guzzo & Dickson, 1996); (2) social decision scheme research (Davis, 1973), which investigates the appropriateness of decision strategies for different task types (e.g., Hackman, Brousseau, & Weiss, 1976; Ladbury & Hinsz, 2009; Laughlin & Ellis, 1986; Shiflett, 1972; Stone, 1971) ; and (3) research that tries to connect social decision scheme literature with social influence literature by studying the interplay between member resources, social decision schemes and interaction processes, and their relation to team performance (e.g., Bottger & Yetton, 1988; Einhorn et al., 1977; Yetton & Bottger, 1982).

The study also contributes to research on fast and frugal heuristics. It revealed that an overall highly predictive accuracy of the recognition heuristic for individual choices (here 83%) does not necessarily lead to a highly predictive accuracy of RBM for group decisions (overall 74%)—but only for those with a high group recognition validity (here .74; the predictive accuracy of RBM in RBM groups was 85%). Groups may be more sensitive in detecting how strongly someone believes in the recognition argument. While on an individual level there is little else one can do than to apply the recognition heuristic, on the group level there is the question of whether to mention (non-)recognition or not. Furthermore, in contrast to individuals, groups have the chance to gather information about both objects even if one object is not recognized by some, while an individual does not have access to information about the unrecognized object. Thus, on the basis of the information collected during the group discussion, it might be more reasonable and successful to decide against the previously unrecognized object if more arguments (or people) are in favor of it. Still, we argue that studying how people perform the same task as individuals and in groups can help to illuminate how heuristic strategies are used (see also Reimer, Hoffrage, & Katsikopoulos, 2007). Of more theoretical importance and relevant for the debate on which role recognition plays in decision making (e.g., Hilbig, 2010) is our finding that recognition was verbalized and used as an argument during discussion and perceived as a highly valid piece of information.

The fact that we found a number of groups predominantly relying on their members using only recognition information (RBM groups) and that these groups performed very well (and even better than KBM groups) may be seen as another piece of evidence for the less-is-more effect on the group level (Gigerenzer et al., 1999; Reimer & Katsikopoulos, 2004). This result is of great practical importance, given the often observed difficulty that groups face when pooling and integrating many pieces of information (Stasser & Titus, 2003; Tindale & Sheffey, 2002; Winquist & Larson, 1998). It implies that the performance of a group is not necessarily raised only by raising the quantity of information exchanged, which was the goal of much previous research (e.g., Frey, Schulz-Hardt, & Stahlberg, 1996; Larson et al., 1994; Parks & Cowlin, 1996; Stasser, Taylor, & Hanna, 1989; Wittenbaum & Stasser, 1996; for a discussion see Reimer & Hoffrage, 2003). Rather, the adaptive selection of group decision strategies determines the success of a group. This means that a group has to select a strategy that fits to the structure of the task environment *and* to the features and composition of the group members (cf. Bottger & Yetton, 1988; Hill, 1982). Counterintuitively, to be successful therefore can sometimes require betting on less knowledgeable members.

Acknowledgments

This research was funded by the Max Planck Institute for Human Development, Berlin, Germany, and supported by the Humboldt University Berlin. We would like to thank Susanne Abele, Konstantinos Katsikopoulos, Ulrich Klocke, Thorsten Pachur, Wolfgang Scholl, Gary Stasser, Michaela Turß, Katharina Wagner, and Claudia Walther for many helpful comments. We thank Ann-Christine Massenberg and Birgit Silberhorn for their help with coding the discussions, and Anita Todd for editing the manuscript.

Chapter 4

The influence of task difficulty on advice-taking behavior: An ecological rationality perspective

Kämmer, J. E., Moussaïd, M., Pipergias Analytis, P., & Neth, H.

Abstract

Many decisions are made in a social context, for example, under the advice of another person. Integrating the opinions of other people can boost one's accuracy. Here, we investigated on the environmental circumstances under which two prominent strategies—averaging and choosing—are more or less effective and adaptive, and how people employed them. We report about an experiment in which participants (N = 90) gave initial estimates for general knowledge questions of three different levels of difficulty, and then received the estimate and confidence rating from another person (advice) before giving a second estimate. We found that items of different difficulty levels exhibited different statistical properties, thus constituting different environmental structures. These affected the (potential) performance of distinctive strategies (such as averaging and choosing), and the way people integrated advice, as suggested by the framework of ecological rationality.

Introduction

Uncertainty shapes our lives; thus, as a coping mechanism, people often turn to others for advice. Dependent on the given situation, advice taking may have a small-scale or a large-scale impact. Consider, for example, that you need to estimate the prospects of an investment. After seeking the advice of a consultant for this purpose you will have to decide how to integrate the advice with your own opinion. Would it be wise to follow the advice or, in case, it differs from your own guess, to go somewhere in between your guess and that of the advisor? Think of a different, more mundane, example: a game evening with your friends where you consult your game partner to give an answer in a quiz with general-knowledge questions. Whereas wrongly discounting or overweighting the advice may have little consequences in such a mundane situation, it may cost or win you a million in the quiz show "Who wants to be a millionaire?" What is the best way to use advice in these two game situations? Contemplate on the possibility that the questions in these two game situations differ in their difficulty, whereby the one in the quiz show might be much more difficult than the other. Given such a difference, should you—and will you—react with different strategies to make use of the advice in the two situations?

The goal of the current study is to investigate the impact of task difficulty on advice-taking behavior, from a prescriptive as well as descriptive perspective. In other words, it studies whether different task difficulties require distinct advice-taking behaviors and whether people react accordingly. We focus here on the judge-advisor system (JAS; Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001) that engages in quantitative estimations: Judges give an initial estimate, receive advice in form of an estimate of another person and then revise their first estimate. How people arrive at their initial estimates in the first place is not the topic of this paper but has been studied extensively elsewhere (see e.g., QuickEst heuristic, cf. Hertwig, Hoffrage, & Martignon, 1999).

The underlying assumptions here are (1) that there exist multiple advice taking strategies and (2) that no single advice taking strategy will perform best in every task environment; rather distinct environmental structures favor different strategies (cf. Soll & Larrick, 2009). In estimation tasks, people were found to primarily rely on two distinct advice taking strategies: choosing and averaging (Soll & Larrick, 2009). Averaging refers to taking the mean of (two) continuous estimates, whereas choosing refers to either staying with the own initial guess (choose the self) or adopting the advisor's guess (choose the other).

Our two main assumptions find their parallels in research on decision making (e.g., Gigerenzer, Todd, and the ABC Research Group, 1999; see below), namely in the ecological rationality framework. Its core idea is that behavior is a function of the (cognitive) resources at hand *and* the environmental structure, and its success depends on the fit between the two. Taking the perspective of ecological rationality to study advice-taking behavior, our main line of argument is thus (1) that tasks of different difficulty levels will constitute distinctive environments that (2) shape the effectiveness of alternative strategies and (3) also the behavior of judges. In the following, we will provide evidence for these three assumptions and also investigate whether people behave in an adaptive way, that is, select the most appropriate strategy. We start with briefly introducing the framework of ecological rationality in more detail.

Ecological Rationality

Central to the concept of ecological rationality is the idea that the human cognition is an adaptation to the structure of the environment and that people possess a variety of cognitive strategies to adaptively deal with problems in distinctive environments and with different capacities at hand (e.g., Simon, 1956). The analogy of an adaptive toolbox (Gigerenzer et al., 1999) provides an image of such a repertoire of strategies, of which none is an all-purpose tool but rather developed to fit to specific environmental structures. The question of whether, when and how people select appropriate strategies when being confronted with different task demands has been investigated in depth, and many scholars have provided evidence for adaptive strategy: For example, it has been shown that task characteristics such as costs of information search or time pressure influence decision making (e.g., Bröder, 2003; Christensen-Szalanski, 1978, 1980; Payne, Bettman, & Johnson, 1988, 1993). Moreover, environment characteristics such as the dispersion of cue validities and information redundancy have been found to influence decision making in a systematic way (e.g., Dieckmann & Rieskamp; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2006).

Whereas the study of these relationships has been the focus of numerous studies on individual judgment and decision making (for an overview see Gigerenzer & Gaissmaier, 2011; Todd, Gigerenzer, & the ABC Research Group, 2012), little attention has been directed to social contexts such as group decision making or advice taking (for a few examples see Biele & Rieskamp, in press; Kämmer, Gaissmaier, Reimer, & Schermuly, 2012; Reimer & Katsikopoulos, 2004; Soll & Larrick, 2009). With this study, we aim to draw more attention to the question of whether people appropriately use advice in different environments, and by this build on recent work by Soll and Larrick (2009).

Adaptive Advice-Taking Behavior

Previous research has concluded that people engage in one general strategy when integrating advice, namely adjusting their own guesses by 30% towards the advice (termed the self/other effect; Harvey & Fischer, 1997; Yaniv, 2004; Yaniv & Kleinberger, 2000). Soll and Larrick (2009), in contrast, revealed that people mainly engage in two distinct strategies, choosing and averaging. The differences occurred due to the use of different methods: Whereas previous research had used aggregate analyses, Soll and Larrick (2009) conducted individual-based analyses. Thus, averaging and choosing may be considered as two tools in the toolbox of advice-taking behavior (cf. Hertwig & Herzog, 2009). Their effectiveness is determined by the structure of the environment (Soll & Larrick, 2009), whereby in advice situations the environment of a judge is constituted by the (distribution of) knowledge of all possible advisors. In the typical JAS of just one judge and one advisor, one person constitutes the "environment" of another person. Soll and Larrick (2009) identified two relevant environmental characteristics that shape the effectiveness of different advice taking strategies and summarized them in their PAR model. In the following, we will describe the PAR model and its predictions concerning the relative performance of averaging and choosing and highlight some links to related research.

The PAR model (Soll & Larrick, 2009)

According to the PAR model, three factors determine the relative accuracy of averaging versus choosing: the probability of detecting the better judge (P), the relative *accuracy* of two judges (A), and the *redundancy* of errors (R) (cf. Soll & Larrick, 2009). If a judge receives advice on a number of questions, probability p is defined as the probability of detecting the more accurate judge for the entire set of questions. It depends on the availability and appropriate use of good cues to expertise und thus describes a feature of the judge. The latter two factors, in contrast, describe environmental conditions. To illustrate, Figure 1 depicts an example. Assume the true answer to an estimation task is 500, and two judges estimate it to be 450 (x_1) and 600 (x_2), respectively, thus erring by 50 (e_1) and 100 (e_2). The difference in their errors is then captured by the accuracy ratio A, which is the ratio between the (mean) absolute deviations from the truth (A = e_2 / e_1), indicating how different judges are in terms of their information or skills. In our example, the worse judge is twice as inaccurate as the better judge (A = 100 / 50 = 2).

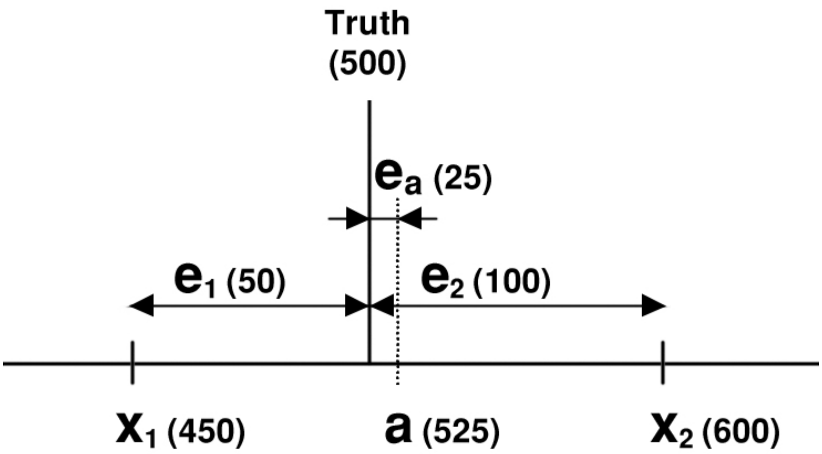


Figure 1. Illustration of an estimation task. Let the true answer be 500 (Truth) and x_1 and x_2 the estimates of two persons. Their respective errors are computed as the absolute distance to the truth (e_1 = |Truth – x_1| and e_2 = |Truth – x_2|). Averaging the two estimates [a = (e_1 + e_2)/ 2] yields a smaller error (e_a = 25) than choosing either of the two judges (e_1 = 50, e_2 = 100).

Redundancy is operationalized by the bracketing rate Br. In Figure 1, each of the two estimates lies on one side of the truth, thus bracketing the truth. In general, the bracketing rate is “defined as the proportion of questions for which the true answer lies strictly between the estimates of the two judges” (Soll & Larrick, 2009, p.784). Low bracketing rates indicate a larger correlation between the judges’ errors and a common bias towards one side of the truth (meaning that the majority of people tend to either over- or underestimate the truth). Low bracketing rates may be, for example, the result of judge and advisor having a similar background, experience or exposure to the media. How do these three factors influence the performance of choosing and averaging?

Choosing

As mentioned before, choosing encompasses two distinct processes, namely that judges stay either with their own initial guesses or switch to the advisor's guess. Although these two constitute two psychologically different processes, concerning their theoretical performance they can be regarded as one strategy if one assumes that judge and advisor come from the same population. A related strategy in group decision making may be seen in "follow-the-expert" (Einhorn, Hogarth, & Klempner, 1977) and in cognitive psychological research the take-the-best heuristic (Gigerenzer & Goldstein, 1996). They have the logic in common to identify the better or best (judge, group member or cue) among the available ones and then to follow this one while ignoring the remaining ones.

The PAR model predicts that large differences between judges (high A) favor choosing because one judge is clearly better than the other. Similarly, take-the-best performs well if cue validities are highly dispersed (Dieckmann & Rieskamp, 2007; Hogarth & Karelaia, 2007). Moreover, for choosing to outperform averaging, a good cue to recognize these differences must be available (high p) and the judges' errors may be correlated. Again, a parallel exists to group decision making, where only large differences in expertise between group members favour a best-member rule, whereas small or unknown differences in expertise favour a simple majority rule (Einhorn et al., 1977). A common finding from small group research also highlights the importance of the third precondition for choosing to outperform averaging: If the true answer cannot be demonstrated or expertise is not recognized by the members of a group, the group performs below the level of the best member and is better off by following the majority (e.g., Bonner, 2004; Henry, 1995; Libby, Trotman, & Zimmer, 1987).

Averaging

Averaging simply refers to taking the average of the continuous estimates of a judge and an advisor to form a revised estimate. In Figure 1, the average estimate is denoted a, and its corresponding error is e_a. A corresponding strategy in decision making is unit weighting of cues (Dawes, 1979; Einhorn & Hogarth, 1975), whereby (usually more than two) cues receive the same weight (.5 in the case of two cues) when being integrated.

Research on the wisdom of crowds is a prominent example of pointing to the advantage of averaging multiple opinions (Galton, 1907; Surowiecki, 2004). The power of averaging has statistical roots. To start with, an estimate can be decomposed into a true value and an error term (Novick, 1966). Averaging estimates then can cancel out random errors and reduce systematic errors made by individual judges (Page, 2007; Soll, 1999; Surowiecki, 2004; Yaniv, 2004). In order for averaging to yield a smaller error than randomly choosing one single estimate, the estimates have to lie on both sides of the truth, that is, to bracket the truth (Larrick & Soll, 2006; see also Herzog & Hertwig, 2009). Hence, somewhat counterintuitively, even advice from non-experts can turn out to be useful, when it lies on the other side of the truth from one's own estimate (e.g., Yaniv & Kleinberger, 2000). Figure 1 illustrates such a case where averaging yields a smaller error than either of the two initial estimates because the two estimates x_1 and x_2 bracket the truth.

More generally, pooling the opinions of others, for example, by applying a majority rule (Hastie & Kameda, 2005) or averaging (multiple) quantitative guesses (Galton, 1907; Hogarth, 1978) often outperforms strategies that take into account only single opinions, and in the extreme, even single expert opinions (e.g. Krause, James, Faria, Ruxton, & Krause, 2011; Sjörberg, 2009). Independence of judgments (i.e., a high bracketing rate), however, is an essential precondition for averaging to perform well (Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Surowiecki, 2004). To illustrate, Herzog and Hertwig (2009) showed that the error of the second (here: advisor's) estimate can even be three times as large as the judge's initial estimate and averaging would still yield a better result than choosing the self—under the condition that the two estimates bracket the truth.

In sum, the model predicts that averaging is superior when the more accurate judge can be identified only with a low probability, when accuracy differences are low, and when errors are uncorrelated, that is, bracket the truth. To make use of the high potential performance of choosing under conditions of high accuracy differences and redundant errors, a sufficiently high probability *p* is necessary. Lower probabilities of identifying the better judge result in a larger range of situations in which averaging outperforms choosing (see Figure 3 in Soll & Larrick, 2009, p. 790). In other words, the relative advantage of choosing over averaging is a function of p. In a series of four experiments, Soll and Larrick (2009) provided first evidence for the predicted impact of the three conditions accuracy ratio, bracketing rate and probability on the potential performance of choosing and averaging. In the following, we will apply the PAR model to study the effect of task difficulty on advice taking, which has not been done before.

Environmental Properties of Tasks with Varying Difficulty

First, we argue that tasks of different difficulty levels constitute different environments, that is, are accompanied by different environmental characteristics. Task difficulty is typically measured as the proportion of correct answers (e.g., Lichtenstein & Fischhoff, 1977). Another possibility (which we employed) is to measure the *perceived* level of difficulty so to have a measure that is independent of accuracy and thus allows for investigating their interrelationships (cf. Lorenz et al., 2011).

With increasing difficulty, the dispersion of answers in the distribution and the collective error increase (Kämmer, Pipergias Analytis, Neth, & Moussaïd, 2012). Figure 2 shows two schematic distributions of answers, one in an (idealized) easy task environment on the left and one in a difficult task environment on the right.

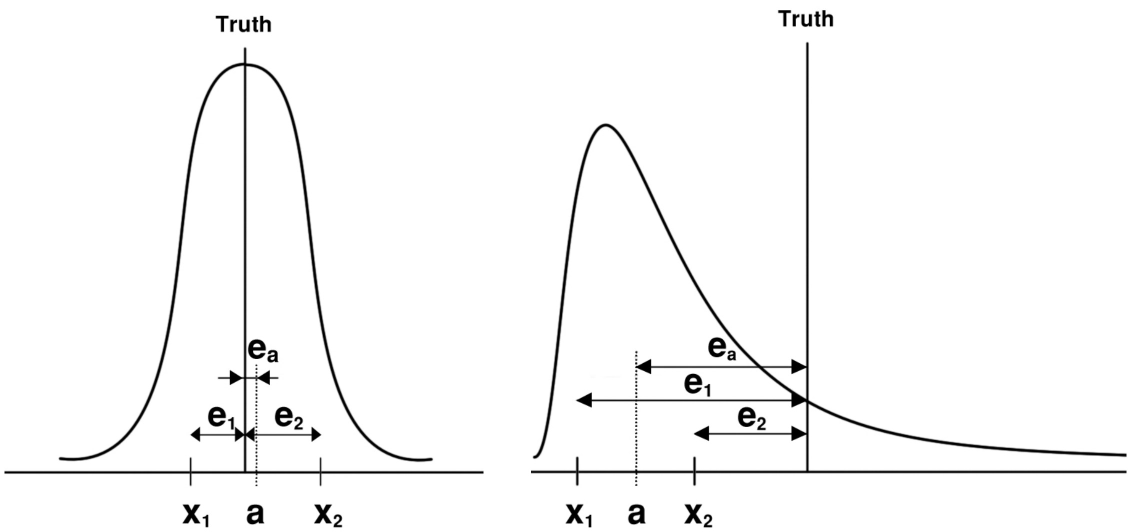


Figure 2. Two schematic illustrations of an easy task environment (left) and a difficult task environment (right).

For easy tasks, the variance of estimates is rather low and estimates lie rather symmetrically around the truth, leading to a low collective error (as illustrated by the left panel in Figure 2). For difficult tasks, in contrast, estimates become more widespread and a systematic distortion to one side of the truth becomes more likely (in the right panel of Figure 2: an underestimation of the truth can be seen). The bracketing rate reflects such distortions and will thus be lower for difficult than for easy tasks. The accuracy ratio is related to variance and thus will also change with changing difficulty:

Hypothesis 1: The accuracy ratio will increase with increasing difficulty.
Hypothesis 2: The bracketing rate will decrease with increasing difficulty.

The constellation under low difficulty will thus constitute environmental conditions that favor averaging, whereas the constellation under high difficulty will favor choosing. However, as pointed out before, the relative accuracy of the two is determined by the third variable in the PAR model, probability of detecting the better judge. In a JAS, we consider confidence being a potential cue for detection. In fact, confidence is a frequent and intuitive cue to judge the accuracy of a source (Sniezek & Van Swol, 2001). Confidence is seen as the strength with which a person believes in her estimate (Peterson & Pitz, 1988; Zarnoth & Sniezek, 1997). Informing people about the advisor's confidence is supposed to help them accurately identify the level of expertise of the advisor (e.g., Sniezek & Buckley, 1995; Sniezek & van Swol, 2001; Soll & Larrick, 2009) and thus weight advice accordingly (Sniezek & Henry, 1989). Obviously, confidence is only beneficial to the extent that it is correlated with accuracy (Yaniv, 1997), which is however not always the case (Savadori, Van Swol, & Sniezek, 2001; but see Sniezek & van Swol, 2001; van Swol & Sniezek, 2005). Difficult tasks are usually accompanied by a greater uncertainty of people and thus a lower confidence in the accuracy of their guesses (e.g., Gino & Moore, 2007). As there is a larger variance in guesses at the same time, we expect the correlation between confidence and accuracy to be lower for difficult tasks than for easy ones.

Hypothesis 3: The probability of detecting the better judge will decrease with increasing difficulty.

We ground this prediction also on the self-consistency model (SCM) of subjective confidence proposed by Koriat (2012a). Koriat (2012a) observed that in choice tasks, the confidence-accuracy relationship is rather a by-product of the consistency-accuracy relationship: “It is positive because the answers that are consistently chosen are generally correct, but negative when the wrong answers tend to be favored” (p.80). Applied to our setting, we may thus expect that confidence will be a good indicator of accuracy for easy tasks because the majority is correct, whereas confidence becomes a bad indicator for difficult tasks because the majority is incorrect, and confidence rather mirrors the consistency of answers than their accuracy in the first place.

Thus, when a good cue to accuracy is most important (i.e., for difficult items), the confidence cue will probably be also least valid, leading to the prediction that choosing will not outperform averaging on difficult tasks because the best judge will not be identified with sufficiently high probability. Averaging will thus also perform better for difficult items (unless p is high) as it makes sure that at least some weight is put on the best judge (cf. Soll & Larrick, 2009). This prediction will hold for judges even when they are better than the advisor under the condition that their estimates bracket the truth (Herzog & Hertwig, 2009). In case of no bracketing, the worse judge will still benefit from averaging; however, the better judge will not.

To sum up the prescriptive predictions, we expect that easy tasks are accompanied by a low accuracy ratio and high bracketing rate leading to good performance of averaging. However, owing to the good cue to expertise, choosing will also perform well. In fact, this constellation might lead to small differences in theoretical performance between the two strategies. For intermediate tasks we expect intermediate levels and for difficult tasks we expect high levels of accuracy differences and lower rates of bracketing. In principle, this constellation constitutes good preconditions for choosing to perform well. Due to the unreliable cue confidence that prevents from telling who is better, however, we expect that averaging will also perform better on difficult tasks given some bracketing.

Hypothesis 4: From a prescriptive point of view, the potential benefit of choosing over averaging will increase with increasing difficulty. Due to a decreasing probability of detecting the better judge with increasing difficulty, however, averaging will outperform choosing also on difficult tasks.

Influence of Task Difficulty on Advice-Taking Behavior

By varying task difficulty, we systematically study a factor that is usually only implicitly varied between experiments on advice taking and has received little explicit investigation so far, although it might play an important moderating role (Gino & Moore, 2007). As a case in point, Soll and Larrick (2009) speculated that difficulty might have varied between their four experiments but had not explicitly manipulated it.

What do we know about the influence of task difficulty on advice-taking behavior? The only study, of which we know that directly assessed the effect of task difficulty on advice taking, found a direct as well as an indirect effect of difficulty on advice taking: namely that, "difficulty increases advice taking by reducing confidence," and beyond, namely "even after controlling for expressed confidence" (Gino & Moore, 2007, p. 32). In this study, the task was to estimate the weight of people from pictures, and difficulty was varied by either blurring the picture or not.

High difficulty is accompanied by low confidence and a lack of knowledge or increased uncertainty (Gino & Moore, 2007). Studies that did not directly manipulate task difficulty but recorded confidence and knowledge found that low confidence as well as a lack of knowledge made people more receptive to advice (e.g., Gino, Brooks, & Schweitzer, 2012; Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000; Yaniv, 2004). In other words, the more knowledgeable people were, the more they were found to discount advice. Drawing on previous studies and from a descriptive perspective, we would thus expect to find a larger degree of advice integration when people face difficult tasks and more discounting when they face easy tasks. As previous findings are, however, mainly based on aggregate analyses and do not specify advice taking strategies any further, more precise predictions about when which specific strategies is used are speculative. Given that averaging was found to be a prominent strategy (Soll & Larrick, 2009), we suspect to find a larger frequency of averaging in difficult tasks.

Hypothesis 5: People will be more prone to integrate advice under high levels of task difficulty than under low levels and probably engage more in averaging.

Overview

Many decisions are made in a social context, for example, under the advice of another person (e.g., Bonaccio & Dalal, 2006) or of multiple other people (e.g., Yaniv & Milyavsky, 2007). The present study examines the question of how advice taking strategies are influenced by environments with varying difficulty from (1) a normative and (2) a descriptive perspective. We study these questions taking the example of quantitative general knowledge estimation tasks, where participants had to give an initial estimate, received advice and had to give a second, revised estimate. We confronted participants with items of three different difficulty levels, and studied the effect of the different environments on the theoretical performance (Which strategies would lead to the best revised estimates?) and on the actual use of strategies (Which advice taking strategies can be observed?), and the fit between the two.

Method

Participants

Participants included 90 people (45 females, M_{age} = 25.21 years, *SD* = 4.24), of whom 61.2% indicated being students. Participants took on average 40 minutes to complete the experimental task and received €10.36 (*SD* = 1.3; €1 = $1.25 at that time).

Design and Procedure

Difficulty was varied between subjects (3 levels: easy, intermediate, and difficult). Upon arrival, participants answered some demographic questions, completed a practice trial and then worked on the experimental task. Participants received rewards in addition to a show up fee of €5 so to be motivated to answer all estimation questions as accurately as possible. Possible rewards were 3, 2, or 1 point for each estimate that fell into the 10%, 20% or 40% intervals around the true value, respectively (1 point = €0.12).

Experimental Task

The experimental task was an estimation task comprising 24 general knowledge questions covering geography, sports, and societal topics. All correct answers consisted of three digits, that is, they ranged from 100 to 999, which, however, was not known to participants. The purpose of this limitation was to reduce estimates to a range of magnitude in which numbers are represented in a fairly homogeneous fashion (see, e.g., Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004).

Participants were first asked to give an initial estimate to all 24 items and to indicate their confidence on a six-point scale (1 *very unconfident* to 6 *very confident*). Participants were then informed about the estimate and confidence rating of another person, who was randomly drawn from the sample of the prestudies (see below). They were then asked to give a second estimate for eight items of the complete set, belonging to one difficulty level, and again to indicate their confidence.

The time for every estimation question was limited to 30 seconds. If time elapsed before an estimate was entered, it counted as a missing value (which only occurred in 0.3% of all trials). Feedback on the accuracy of answers was only provided at the very end of the experiment. After the experiment, participants were paid, thanked, released, and informed about the correct answers.

Item selection

In order to create a sample of 24 general knowledge questions that are commonly perceived as easy, intermediate and difficult, we conducted two prestudies, in which we collected the intuitive estimates and perceived difficulty of a total of 111 participants (study 1: N = 52, M_{age} = 27 years, *SD* = 9, 50% females; study 2: N = 59, M_{age} = 33 years, *SD* = 11, 56 % females) for an initial set of 48 questions (30 in each study with 12 overlapping questions). Participants of the prestudies did not take part in the main study. They received a flat fee of €8 for an approximately 40-minutes lasting experiment. For each estimation question, participants were asked to give their intuitive estimates and to indicate their felt confidence (1 *very unconfident* to 6 *very confident*). Moreover, perceived difficulty was measured with three items:

1. “Please imagine 100 peers that have answered the same question. How easy or difficult do you think the question was for most of these people?” (1 *very easy* to 6 *very difficult*)
2. “How many of these 100 people know the exact answer?” (0 to 100)
3. “How many people’s estimates differed extremely from the true value, that is, their estimate is equal to the true value plus a deviation of more than 30%?” (0 to 100; reverse coded)

To better compare these measures, we then normalized the confidence and the three difficulty measures to the same scale (0 *easy* to 1 *difficult*). As expected, the four normalized values (as well as the raw values) were highly correlated (Spearman’s rho = .848 on average).

For the question selection, we computed the means for each question on the four scales and ranked the 48 questions for each scale. Each question now had four ranks between 1 (*easy*) and 48 (*difficult*). Based on these four rankings, we then selected eight items that appeared four times between the ranking positions 1 and 14 (*easy* items), 15 and 34 (*intermediate*) and 35 and 48 (*difficult*), respectively. We thus used the convergent ratings from four different difficulty measures to select commonly perceived easy, intermediate and difficult items, independent of the objective accuracy of participants. The final list of items together with the average perceived difficulty can be seen in Table D.1 in Appendix D. Moreover, screenshots of the experimental paradigm can be seen in Figure D.1 in Appendix D.

Advice selection

The 111 first estimates collected in the prestudies constituted the pool of possible advice. For every question, the estimates and corresponding confidence ratings of seven participants out of this pool were randomly drawn with the restriction that the selected sample’s mean and standard deviation would not deviate more than 25% from the overall mean and standard deviation, respectively. This representative sampling was meant to ensure ecologically valid advice, which participants might have encountered in reality. For every participant, one estimate plus the corresponding confidence was randomly drawn from this selected sample.

Dependent measures

In total, we collected 2,160 first estimates (24 items × 90 participants) and 240 second estimates per difficulty level (8 items × 30 participants; 3 difficulty levels). We first conducted an outlier analyses on the first and second estimates. Using Tukey's method (1977), extreme values that were located at least 1.5 interquartile ranges (IQR) below the 25th percentile (Q1) or above the 75th percentile (Q3) were regarded as outliers (Q1 – 1.5IQR, Q3+1.5IQR). This widely-used method has the advantage of not being limited to normally distributed data and of being resistant to extreme values (Seo, 2006). 8.7% of the first estimates and 8.4% of the second estimates were identified as outliers and eliminated.

Both first and second estimates were standardized at the true value ($x_s = x / truth$, where $x = estimate$) so that the estimates became comparable across items. To calculate the error, we further subtracted 1 from the standardized value and took the absolute ($e = abs[1 - x_s]$), so that the truth equaled zero now; large values represented large errors. Subscripts 1 and 2 denote the first or second estimate (e.g., e_1 for error of estimate 1).

To measure the extent to which judges integrated the advice versus kept their own initial estimate, we computed the weight on self (*ws*; Soll & Larrick, 2009): $ws = \text{abs}(x_2 - a) / \text{abs}(x_1 - a)$, where x_1 is the initial estimate, x_2 the revised estimate and a the advice. This measure usually takes values between 0 (indicating that the advice was adopted) and 1 (when the revised estimate equals the initial estimate, that is, the judge chose the self). Occasionally, revised estimates lie outside the range of the judge's initial and the advisor's estimates (cf. Bonaccio & Dalal, 2006). In our study, 9.44% of *ws* values were larger than 1 (and thus occurred slightly more often than the usually observed 5%, cf. Bonaccio & Dalal, 2006, p.141; none was below 0). In 27 cases (4.34%), the judge's initial estimate was equal to the advice, so that *ws* could not be computed (because the denominator becomes 0).

Results

We structure the results part along our three goals: We first establish whether our different difficulty levels are accompanied by some measurable statistical properties in the expected directions. We then analyze the impact of difficulty on the theoretical performance of advice taking strategies. Lastly, we study how participants behaved in the three difficulty environments and compare this with the achievable performance.

Do Environments Differ in Some Relevant Statistical Properties?

Table 1 contains the statistical properties of the three item sets belonging to the three difficulty levels. Recall that each item set consisted of eight items.

To start with and as a kind of manipulation check, we analyzed the confidence ratings and average error. As can be seen in Table 1, the average confidence levels decreased with increasing difficulty as predefined by our prestudies (for a comparison with the confidence ratings of the prestudies see Table D.1 in Appendix D). Also, the average individual error and variance of estimates increased with increasing difficulty. To illustrate, Figure 3 depicts the distribution of standardized answers for one sample item per difficulty level, respectively.

Table 1

Statistical properties of item sets per difficulty level.

Statistic	Item sets easy	intermediate	difficult
Mean confidence (SD)[1]	3.48 (0.88)	2.54 (0.93)	2.00 (0.69)
Mean distance between confidence ratings of two judges (SD)[2]	1.43 (1.12)	1.45 (1.17)	1.13 (1.04)
Mean individual error of initial estimate (SD)[1]	0.29 (0.29)	0.60 (0.48)	1.06 (2.06)
Mean individual error of revised estimate (SD)[3]	0.23 (0.26)	0.46 (0.35)	0.49 (0.68)
Mean intraindividual correlation between error and confidence (Spearman's *rho*)[1]	- .41 (.33)	- .24 (.43)	- .01 (.43)
% of cases in which the more confident judge was also the more accurate one (out of cases in which the two estimates differed in accuracy and confidence)[2]	63.57%	58.93%	53.39%
Mean lognormal distribution parameter (scale parameter)[1]	0.58 (0.40)	1.24 (0.34)	1.47 (0.58)
Mean accuracy ratio A[2]	1.56	1.42	2.05
Mean bracketing rate Br[2]	0.41	0.26	0.34
Mean difference between errors of two judges[2]	0.11	0.19	0.70

[1] Values were calculated for $N = 90$ participants.

[2] Values were calculated on a complete pairing of the standardized initial estimates (x_s) of $N = 90$ participants (resulting in up to 4005 unique pairs).

[3] Values were calculated for $n = 30$ participants per difficulty level, as difficulty was varied between-subjects.

In Figure 3, it can also be seen that most initial estimates followed a lognormal distribution, which was skewed to the right (significant for 33% of them: for 50% of easy items, and for 25% of intermediate and difficult items, respectively). In Table 1, we report the corresponding average lognormal distribution parameters, that is, the standard deviation of the estimates' logarithm (scale parameter). It increases with increasing difficulty, thus bmaking it a good indicator of the difficulty level too.

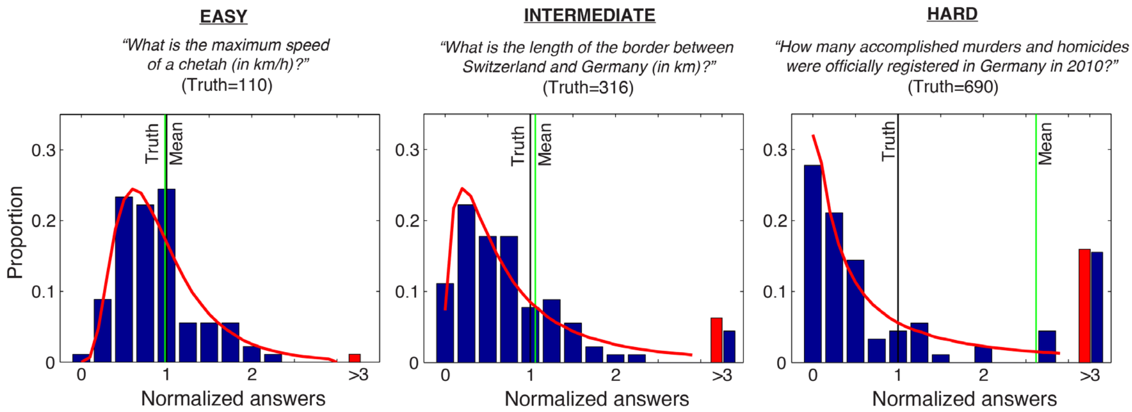


Figure 3. Initial distributions of estimates (x_{s1}, including outliers) for one representative sample item of each difficulty level, respectively. The normalized answers correspond to the estimates divided by the true value. The correct value, therefore, corresponds to 1. The red curves show the best fit of a lognormal distribution, and the red bars indicate the theoretical proportion of outliers having a normalized answer higher than 3.

According to the PAR model of Soll and Larrick (2009), two aspects describe the environment and allow for predictions about the relative theoretical performance of choosing and averaging: the accuracy ratio and the redundancy between advisor and judge, measuring the diversity and distribution of errors. Both were calculated for each possible pair of the first estimates on the eight questions per difficulty level and then averaged. For the observed values of the accuracy ratio and bracketing rate in the advice condition see Table 2.

Accuracy ratio *A*

Recall that the accuracy ratio *A* is the ratio between the mean errors (e_I) of two judges, higher over lower. As expected, we observed that the accuracy ratio increased from easy to difficult items (thus supporting *Hypothesis 1*, or in short *H1*). On difficult items, for example, the worse judge was on average twice as inaccurate as the better judge. The reason for the rather unexpectedly high ratio for easy items (as compared to intermediate items) is a mathematical one: very small errors, which only appear in the easy environment, are in the denominator when computing the ratio, exaggerating actual differences in the accuracies (e.g., 0.3 / 0.1 = 3 vs. 0.5 / 0.3 = 1.67). We therefore also report the average absolute difference between the errors of two judges, respectively (abs[$e_{1\_judge1} - e_{1\_judge2}$]). Here, the expected steady increase with increasing difficulty is observable.

Bracketing rate *Br*

All in all, bracketing rates were above zero and rather similar to each other, with the largest bracketing rate for easy items and smaller ones for intermediate and difficult ones (*H2*).

Predictions based on *A* and *Br*

Environments with low differences in accuracy among people (low *A*) and high bracketing rates, as it is the case for easy items, favor averaging over choosing (Soll & Larrick, 2009). Due to the high bracketing rate, errors cancel out. On the contrary, environments with large differences in accuracy (high *A*) and a majority of people lying at one side of the truth (small *Br*), as we found it for difficult items, favor choosing over averaging. A third precondition, however, must hold in order for choosing to outperform averaging: People need to have a good cue to identify the judge with the higher accuracy (high probability *p*), since choosing the best over the second best judge makes a big difference due to the large accuracy differences between people.

Probability *p*

In our study, a possible cue to accuracy was provided by the confidence of the advisor. How well was confidence calibrated and indicative for actual accuracy? We report the average intraindividual correlation between confidence ratings and error in Table 1 (note that negative correlations indicate that smaller errors were accompanied by higher confidence ratings). It can be seen that the average correlations were *rho* = -.41 for easy items and *rho* = -.01 for difficult items.[1] The histograms of correlations in Appendix E (Figure E.1) further show that, whereas 88.8% of correlations were

[1] The decrease in correlations is not merely the result of a decrease in the range of confidence ratings, which accompanied the increase in difficulty. When we analyzed the intraindividual correlations between confidence and error for the limited range of confidence 1 to 3 only, we find very similar correlations: *rho* = -.40 (*SD* = .46) for easy items, *rho* = -.20 (*SD* = .45) for intermediate, and *rho* = -.06 (*SD* = .45) for difficult items, indicating a decrease in calibration.

negative in the easy environment, only 69.8% and 44.4% were below zero in the intermediate and difficult environments, respectively.

To get a more intuitive understanding of the validity of the confidence cue, we provide a second measure of the quality of the confidence cue: the percentage of cases in which the more confident judge was also the more accurate one (out of cases in which the two judges differed in accuracy and confidence). It can be seen from Table 1 that for easy items confidence was indicative of accuracy in 63.6% of cases. This proportion decreased with increasing difficulty, however, still being larger than random for difficult items (53.4%). All in all, these results indicate that confidence was quite a valid cue for easy items but not for difficult items on average (supporting *H3*).

Summary

So far, we have constituted that items of different perceived difficulty levels also vary in a number of measurable statistical properties. From a prescriptive point of view, we can thus summarize that averaging should outperform choosing on easy items, whereas the opposite should be the case on difficult items, given a high p. As p plays an important role for actual behavior, however, and turns out to be rather weak for difficult items averaging will be also better in the difficult environment (*H4*).

Is the Performance of Different Strategies Influenced by the Difficulty Level?

In Figure 4 the theoretical performance of choosing and averaging can be seen (in terms of average expected error e of a strategy: large values indicate low performance). Values were calculated from the complete pairing of the first estimates of all participants. To illustrate this procedure, imagine that every participant would receive the first estimate of every other participant once as advice and then always averaged his / her own first estimate with the advice, for example; then one would expect an average error of 0.24 on easy items (the same would be expected for the strategy “randomly choosing” between judge and advisor, because initial estimates are, on average, equally good). Consistently choosing the self or the other would result in a similarly small error of 0.29.

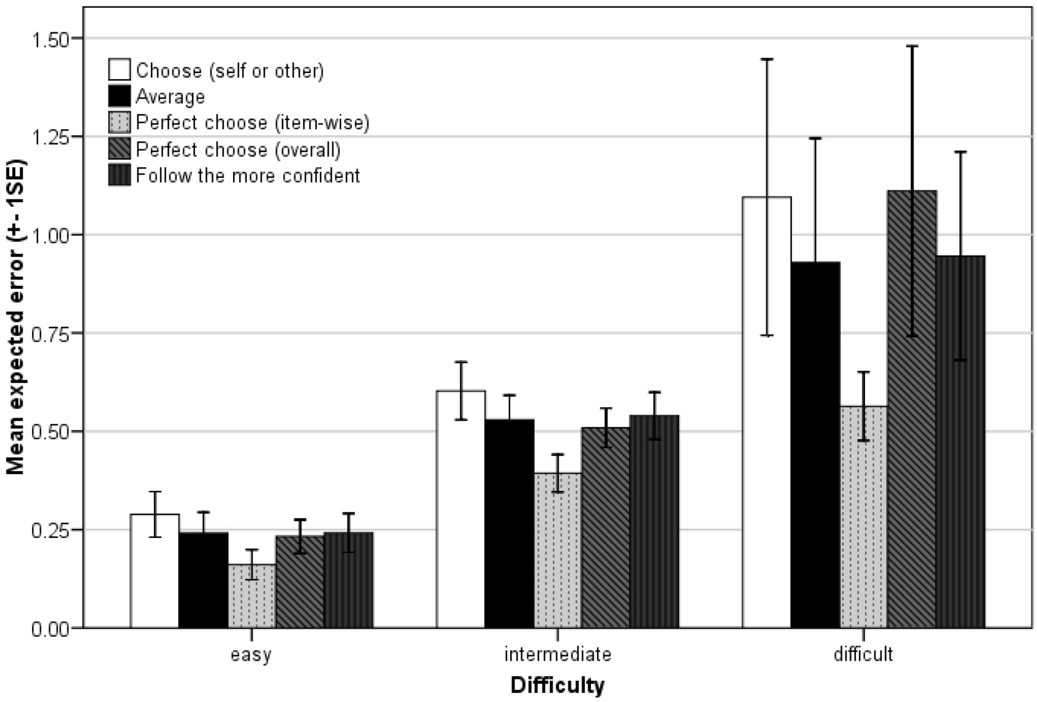


Figure 4. Mean theoretical performance of strategies (average error) per difficulty level. Values are calculated on a complete pairing of the errors of the initial estimates (e_1) of N = 90 participants. Error bars ± 1SE.

The two strategies are compared with the theoretical performance of a benchmark strategy: perfectly choosing the more accurate. Two ways of operationalizing this strategy are possible: (1) item-wise and (2) overall (difficulty level-wise). Measure (1) assumes that people adopt on every single item the better estimate, whereas measure (2) is less strict (and less accurate) as it assumes that people adopt the estimates of the overall better judge (as proposed by Soll & Larrick, 2009; whereby "overall" in our case means being better on the eight items on which people got advice). "Follow the most confident" predicts that the judge adopts the estimate of the more confident judge on each item (cf. Koriat, 2012b) and, in case there is no difference in confidence, averages (percentage of averaging cases was 22.0% for easy items, 23.7% for intermediate, and 30.9% for difficult items).

The three latter strategies implicitly reflect the moderating role of *p* on performance: In case of a perfect cue to identify the better judge, the resulting accuracy corresponds to the item-wise measure. In case of confidence being the cue, the resulting accuracy corresponds to "follow the more confident." If confidence perfectly correlated with accuracy, the two strategies would converge. In case of feedback on the overall performance of the advisor being the cue, the resulting accuracy corresponds to the overall-measure. Note that item-wise perfect choosing and "follow the more confident" are in contrast to choosing and perfect choosing (overall) flexible strategies, in that they do not bet on one judge to be always the best. Averaging and choosing the self or the other do not require any cues in contrast to perfect choosing. Averaging is the only strategy that results in an estimate that is different from the two given ones.

It can be seen in Figure 4 that for easy items the choice of strategies does not have a large impact on the expected accuracy. All strategies yield a low error on average because people's estimates are rather similar. For intermediate and difficult items consistently choosing the self or the other performs worse than averaging, "follow the more confident" and perfect choosing (item-wise). This reveals the discrepancy we expected (*H4*): With increasing difficulty and given a good cue, perfectly choosing the better judge (item-wise) would lead to a better performance than averaging. Taking confidence as a cue, however, becomes less indicative and thus results in a lower performance. Similarly, relying on the overall performance as an indicator of the item-wise performance performs worse because intraindividual accuracy differences increase.

To sum up the prescriptive findings, strategies perform similarly well on easy tasks. With increasing difficulty, the distance between the expected performances of perfect choosing and averaging increases. The lower reliability of cues to expertise (such as of confidence or the overall performance) is reflected in the lower accuracy of strategies that rely on these cues, so that averaging becomes the most efficient strategy for difficult items, not requiring any cues (*H4*).

How Does Difficulty Influence Strategy Use?

We now turn to the third step of our analyses: how people integrated advice in the different conditions. Table 2 contains the observed environmental parameters and errors in round 1 and 2 (rows "all") of the 30 participants in each condition. It also depicts the values for two subgroups, choosers and combiners, to which we will refer later.[2]

---

[2] Note that values slightly differ from the theoretical ones reported in Table 1. The differences occur to the previous section (Table 1) occur because judge and advisor were randomly drawn from the same

Table 2

Average observed values of environmental parameters and accuracy of initial (e_1) and revised estimates (e_2), per environment and subgroup (SD in parentheses).

Env.		n	*A*	*Br*	*rho*	e_1	e_2
	all	**30**	**1.6 (.68)**	**.43 (.19)**	**-.49 (.27)**	**0.30 (0.11)**	**0.24 (0.10)**
easy	choosers	12	1.7 (.90)	.42 (.16)	-.38 (.26)	0.22 (0.09)	0.19 (0.10)
	combiners	18	1.5 (.52)	.44 (.16)	-.56 (.25)	0.36 (0.08)	0.27 (0.09)
	all	**30**	**2.3 (2.1)**	**.39 (.20)**	**-.17 (.43)**	**0.58 (0.14)**	**0.45 (0.13)**
med.	choosers	7	3.2 (4.3)	.29 (.12)	-.38 (.44)	0.52 (0.13)	0.40 (0.13)
	combiners	23	2.0 (1.9)	.42 (.21)	-.10 (.42)	0.60 (0.14)	0.47 (0.13)
	all	**30**	**2.8 (2.1)**	**.32 (.16)**	**-.13 (.34)**	**1.12 (0.76)**	**0.81 (0.36)**
difficult	choosers	3	3.9 (2.3)	.33 (.07)	-.07 (.16)	0.73 (0.25)	1.05 (0.39)
	combiners	27	2.7 (2.1)	.32 (.17)	-.14 (.35)	1.17 (0.78)	0.78 (0.35)

Env = environment, med = intermediate, *A* = accuracy ratio, *Br* = bracketing rate, *rho* = intraindividual correlation between confidence and error

Achieved performance

The aim of advice seeking and taking is to improve judgments. Therefore, we shortly turn to the empirical question whether participants improved their estimate with the help of advice from round 1 to round 2. Column e_1 in Table 2 contains the error of the initial estimate. Column e_2 contains the error of the revised estimate. We conducted an ANOVA with repeated measures with round as within-subjects and difficulty as between-subjects factors and error as dependent variable. It revealed that participants improved from the first to the second round in all difficulty levels, $F_{\text{round}}(1, 629) = 8.472$, $p = .004$, $\eta_p^2 = .01$, and that errors increased with increasing task difficulty, $F_{\text{difficulty}}(1, 629) = 43.333$, $p < .001$, $\eta_p^2 = .12$. Moreover, it can be seen from Table 2 that (similar to the values reported in Table 1) the accuracy ratio *A* increased with

population of 90 participants, whereas in the following section (Table 2) we report results of the advice condition, where the advisor was randomly drawn from the prestudies. The different sample sizes and the fact that advisors from the prestudy turned out to be slightly worse than the judges from the main study were responsible for the slight differences between the results reported in Table 1 and Table 2.

increasing difficulty, whereas the bracketing rate *Br* and intraindividual correlation between error and confidence *rho* decreased.

Weight on self

We first operationalized strategy choice by the weight on self *ws* as proposed by Soll and Larrick (2009). Values of 0 indicate that the estimate of the advisor was adopted for the second estimate (“choose other”), values of 1 indicate that the initial estimate was kept (“choose self”) and values in between indicate some integration of the two estimates, whereby values between .4 and .6 are regarded as averaging. Values larger than 1 indicate that the second estimate fell outside the range of the two initial estimates (“emergent responses”). As *ws* is a continuous measure, we classified each value into one of eight categories (see x-axis of Figure 5). Figure 5 depicts the distributions of the observed categories of *ws* in the three environments (mean values for the continuous measure of *ws* were M_{easy} = 0.73 (*SD* = 1.24), $M_{\text{intermediate}}$ = 0.63 (*SD* = 0.56), $M_{\text{difficult}}$ = 0.70 (*SD* = 0.50)).

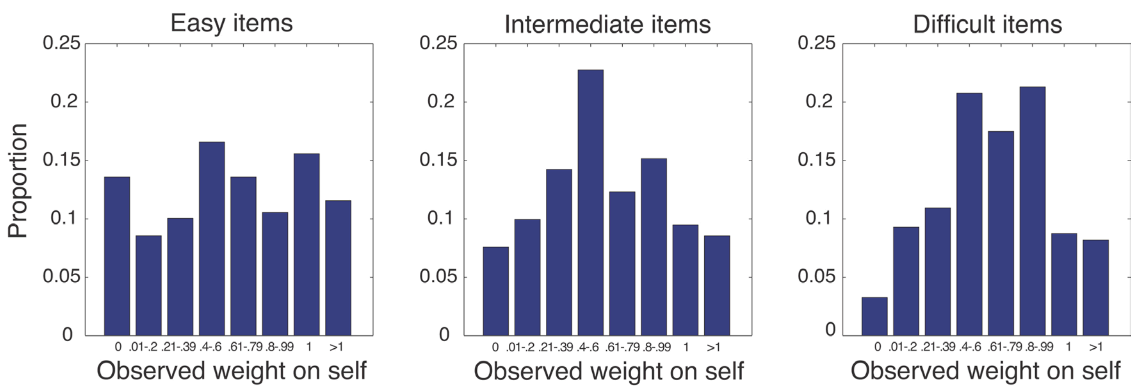


Figure 5. Distribution of weight on self for the three difficulty levels.

Comparisons *within* environments statistically supported what can be intuitively seen in Figure 5 (for better readability, all t-test results are summarized in Table E.1 in Appendix E): paired t-test on the arcsin-transformed proportions of choosing and averaging (Keppel & Wickens, 2004) at Bonferroni’s α = .05 / 3 = .016 revealed the following: In the easy environment, participants chose as often themselves as they chose the advisor and as they averaged. In the intermediate and difficult environments, participants averaged more often than they engaged in choosing (themselves or the advisor). In sum, within-environmental analyses revealed no differences between the modes of choosing and averaging in the easy environment, but more averaging than choosing in the intermediate and difficult environments.

To compare the proportions of the eight categories of *ws between* difficulty levels, we conducted a MANOVA with the (arcsin-transformed) proportion of each category and the difficulty level as factor. No differences between the difficulty levels were revealed for the proportions of choosing the self ($ws = 1$), $F(2, 87) = 1.493$, $p = .23$, $\eta_p^2 = .03$, and averaging, $F(2, 87) = 0.789$, $p = .46$, $\eta_p^2 = .02$. Instead, the adjacent category towards averaging but overweighing the self was observed more often with increasing difficulty ($.80 < ws < .99$ rising from 11% to 25%), $F(2, 87) = 2.990$, $p = .06$, $\eta_p^2 = .064$. Participants adopted the estimate of the advisor in 14% in the easy environment, but only in 8% and 3% in the intermediate and difficult environment, $F(2, 87) = 3.812$, $p = .03$, $\eta_p^2 = .08$. The proportions of occasions in which participants weighted the advisor more than the self without exactly adopting her / his opinion ($.01 < ws < .20$ and $.21 < ws < .39$) were relatively constant over difficulty levels, $F(2, 87) = 0.057$, $p = .95$ $\eta_p^2 = .001$ and $F(2, 87) = 0.874$, $p = .42$, $\eta_p^2 = .02$. Similarly, the proportion of emergent responses was very similar in all three environments (8 to 12%), $F(2, 87) = 0.490$, $p = .61$, $\eta_p^2 = .01$. Overall, the modes choosing and averaging accounted for 58% in the easy, 49% in the intermediate and 41% in the difficult environment, with decreasing proportions of choosing the advisor. The proportion of more mixed strategies (i.e., putting an unequal weight on self and advisor) increased with increasing difficulty, with a tendency to put more weight on the self.

Classification according to *ws*

So far, we have analyzed the distribution of *ws* of all participants together. Next, we test on an individual level whether people accorded with different strategies in the three environments. We therefore classified people according to their predominant strategy choice into choosers [if they chose themselves or the advisor (ws = 0 or 1) in at least half of questions] and combiners [if they chose themselves or the advisor (ws = 0 or 1) in less than half of questions]. The rows “choosers” and “combiners” in Table 2 contain the results for these two subgroups. This classification also allows us to empirically test whether people who differ in their predominant strategy choice also differ in "their" environment. Besides the global differences between environments, each individual person also faces environments that differ on a local level because of his / her individual values of *A, Br* and *p* (cf. Soll & Larrick, 2009). These local differences should influence the performance of strategies in the same way as lined out above.

In all environments, more people were classified as combiners than as choosers, although in the easy environment class sizes were almost equal. This supports the overall results depicted in Figure 5 with equal proportions of choosing and averaging on easy tasks and more mixed strategies on difficult tasks. Differences in the environment characteristic *A* between choosers and combiners were smallest in the easy environment. Higher accuracy ratios could be observed for choosers facing the intermediate and difficult environments, which made choosing in fact more ecologically rational. Choosers and combiners did not differ in their bracketing rates for easy and difficult items, but for intermediate ones, where higher values were found for combiners, which made it again adaptive to average. No systematic difference of the quality of the confidence cue could be observed.

In sum, the descriptive results revealed that people used different strategies for tasks differing in difficulty (*H5*). When confronted with easy tasks, participants did not show an overall preference for one strategy, which is ecological rational since strategies made very similarly accurate predictions. When confronted with intermediate and difficult tasks, more participants averaged and weighted the advice to some extent, though overweighing themselves. This was ecological rational as most participants lacked a good cue to identify the better judge.

In a last step, we compared the observed accuracy with the accuracy that would have been achieved by consistently using one strategy in all trials. We therefore operationalized strategy use by applying the previously introduced strategies on the initial estimate of the judge and the advice to derive the accuracy of the respective predictions. This operationalization has two advantages. First, we are able to compare observed accuracies with theoretical ones and second, we can directly test strategies like "follow the more confident", which is not possible with *ws*. Its disadvantages are that strategies assume the same way of integration is used on every trial, thus rather very inflexible.

Comparing Achieved with Achievable Performance

Figure 6 depicts the observed performance after receiving advice (intuitive revision, which equals e_2) and the achievable performance if consistently applying one of the following seven strategies: randomly choosing between the self and the advisor, choosing the self, choosing the other, averaging, perfect choosing (item-wise) and perfect-choosing (overall), and “follow the more confident.” Choosing the self (ws = 1) and the other (ws = 0) are now differentiated (1) because they constitute two psychologically different strategies and (2) because they can lead to different accuracies if judge and advisor perform differently well in the first place.[3]

It can be seen in Figure 6 that for easy tasks, participants performed worse than the upper benchmark strategy perfect choosing (item-wise), but as well as perfect choose (overall) and “follow the more confident,” and better than the remaining strategies. All differences, however, were relatively small as expected (*H4*).

[3] Figure 6 also reveals that advisors were in our study on average worse than judges on intermediate and difficult items, so that averaging suffered. Averaging would perform better if the advisor was better or equally good as the judge (as can be seen for the easy items and in Figure 4).

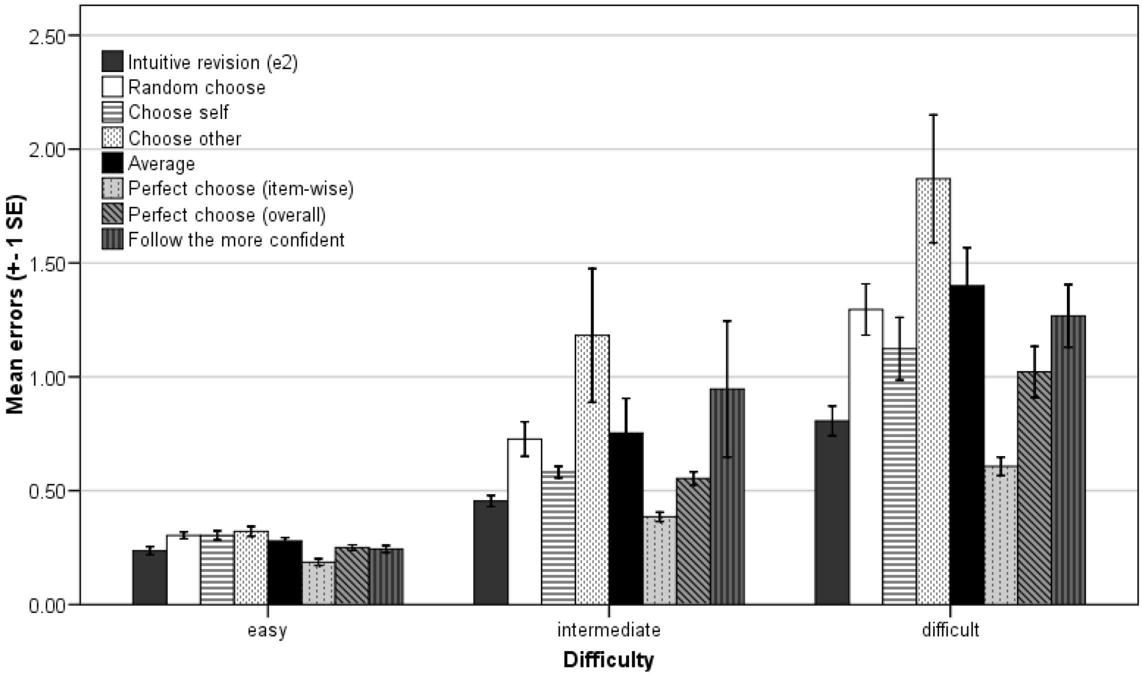


Figure 6. Mean observed and theoretical performance of strategies (average error) per difficulty level. Values are calculated on the errors of the initial estimates of judge and advisor (N = 30 participants per condition). Error bars ± 1SE.

When being confronted with intermediate items, participants again performed better than most of the strategies, except for perfect choosing (item-wise) and equally well as "follow the more competent." On difficult items, participants performed at the level of perfect choosing (overall), worse than perfect choosing (item-wise) but better than the remaining strategies. To sum up, participants performed surprisingly well as compared to the achievable accuracy when consistently following one strategy. The discrepancy between achieved accuracy and perfect choosing (item-wise) can be attributed to the lack of a perfect cue.

Discussion

The main goal of the current study was to investigate the relationship between task difficulty and the performance and use of different advice-taking strategies. We studied this research question from an ecological rationality perspective. In detail, we took three steps: First, based on the PAR model (Soll & Larrick, 2009), we described the statistical properties of item sets of three difficulty levels. As expected, we found a higher accuracy ratio for difficult items than for easy ones (*H1*), leading to a greater potential benefit through choosing the better judge as compared to averaging. The bracketing rate was on a moderate level for all difficulty levels and was slightly lower for difficult than for easy items (*H2*), also favoring choosing. At the same time, however, the confidence – accuracy relation (confidence being the only available cue to expertise) was correlated with difficulty in that it decreased with increasing difficulty, rendering confidence rather a weak cue for the majority of participants facing difficult tasks (*H3*). These constellations suggested that averaging would perform well on all difficulty levels (*H4*).

Second, we analyzed the theoretical performance of choosing and averaging and found that strategies performed very similarly in the easy environment. In the difficult environment, averaging and "follow the more confident" became better than consistently choosing the self or the advisor; but at the same time the distance to the upper benchmark of perfect choosing (item-wise) increased. Thus, perfect choosing would have led to the best performance in principle but already slight deviations (by inferring accuracy on a single item, for example, from confidence or the overall accuracy, which was illustrated by the theoretical performance of perfect choosing (overall) and "follow the more confident") led to underperforming averaging, supporting *H4*.

Third, we analyzed participants' behavior from a descriptive perspective. Average values of weight on self ranged between 63% and 73% in the three conditions and would suggest that people adjusted their initial estimate by around 30%, thus replicating previous work (e.g., Harvey & Fischer, 1997). Analyses of the distributions of the weight on self, however, revealed that these average values were the result of different frequency distributions of choosing and averaging in the three environments: in the easy task environment, choosing the self, the advisor and averaging were equally often used overall. Here, the three modes also accounted for the majority of data (cf. Soll & Larrick, 2009). Individual-level classifications revealed that this result occurred because there were almost equally many participants who mainly either engaged in choosing or averaging, respectively. In contrast, with increasing difficulty the advisor was less often chosen but more averaging and more mixed strategies (putting unequal, but relatively higher, weights on the self than on the advisor) could be observed, mirrored by the increasing number of people classified as combiners. Thus, as in previous studies, people were more likely to integrate advice in difficult tasks (*H5*;

Gino & Moore, 2007). This was ecologically rational as they lacked a good cue to identify the better judge.

Last, we wanted to know if people appropriately adapted their advice-taking strategy to the difficulty level. This question lies at the intersection between the research tradition of advice taking and ecological rationality. We approached this question by comparing the achieved performance with the theoretical performance of seven strategies. Participants were found to perform surprisingly well in comparison with most of the strategies, although they did not reach the level of perfect choosing (item-wise). On easy tasks they performed at the level of perfect choosing (overall) and "follow the more confident" (both converged because confidence was rather a good cue of performance here). On intermediate tasks, they performed at the level of averaging, and on difficult tasks again on the level of perfect choosing (overall). Participants thus outperformed a pure averaging strategy. This may have been caused by (1) the specific constellation in our study as participants met advisors who were slightly worse than them, which caused averaging to suffer, and (2) the inflexibility of most strategies such as averaging, which assume the same integration rule on each trial, whereas perfect choosing (overall), for example, is a more flexible. The distributions of weight on self further suggested that participants used combinatorial strategies with unequal weights (which we did not explicitly tested here as a competing strategy).

Our finding that advice is combined with the judges' own guesses to a larger degree in difficult tasks is consistent with previous studies that found that people were more likely to integrate advice when the task was difficult (Gino & Moore, 2007) or people were uncertain or lacking knowledge (Gino et al., 2012; Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000; Yaniv, 2004). Gino and Moore (2007) explained their finding by taking a person-centered approach arguing that, on easy tasks, people believe themselves to be better than others, whereas on difficult tasks, people think they are worse than others (Kruger, 1999; Moore & Kim, 2003; Windschitl, Kruger, & Simms, 2003), which makes them more willing to accept advice. This argumentation served also as a rationale for the prescription that averaging would have been the optimal strategy in the easy as well as in the difficult task condition, as there should have been as many "participants who believed themselves to be better than others at guessing weights" as there should have been "participants who believed themselves to be worse than others" (Gino & Moore, 2007, p. 27, footnote 3). This rationale is based on the assumption that the beliefs reflect a normally distributed "ability to guess weights." We found, however, that the distribution becomes skewed when the task becomes more difficult, that is, that the actual number of people being good at guessing becomes smaller, whereas the number of people being bad at guessing becomes larger. The result of such a skewed distribution in a difficult task environment would be that most people would correctly believe themselves to be worse than average (for a related argument why most drivers (correctly) say they drive safer than average, see Gigerenzer, 2002, 2004). Taking the mean would thus not necessarily lead to a better result if a good cue to detect the better judge was available, as our prescriptive analyses showed.

The Interplay Between Environment and Behaviour

The insight that much of people's behavior is shaped by the environmental structure (cf. Simon, 1956) had a great impact on research on decision making over the last decades (e.g., Gigerenzer et al., 1999; Todd et al., 2012), but is relatively new to advice taking research (Soll & Larrick, 2009). By finding systematic differences in the environmental structure between tasks of different difficulty levels, our results point to the importance of considering the interplay between the environment and behavior for research on advice taking. Difficulty is usually not directly manipulated or controlled for (for an exception see Gino & Moore, 2007), even though it constitutes an important moderator variable as was shown here. Besides difficulty, other factors may influence the statistical properties of a task environment and thus need to be taken into account when judging which strategy would be "optimal" to use advice. For example, adopting the advice of your Phone-A-Friend lifeline in a quiz show may be most appropriate because this person is an expert in the field (and thus the accuracy ratio is high), whereas averaging the guesses of you and your game partner in the game with friends might lead to the better outcome because you make similarly large errors, which cancel out each other.

We suppose that similar effects of task difficulty can be found in situations in which people elicit a second guess not by asking another person but by asking themselves twice (e.g., with the help of dialectical bootstrapping instructions, Herzog & Hertwig, 2009; see also Müller-Trede, 2011; Vul & Pashler, 2008). Similarly, when people combine estimates of others, the effectiveness of averaging and choosing between them will likely depend on environmental differences caused by varying task difficulty or by other factors that impact *p*, *A*, and *Br* (Soll & Mannes, 2011).

Our detailed analyses of the environmental properties (see Table 1) may be a starting point for future studies that are directed at understanding the interplay between advice-taking behavior and the environmental structure. Varying task difficulty is certainly just one way (though a very intuitive one) to create different environments. The PAR model (Soll & Larrick, 2009) and Table 1 list a number of factors that could possibly be experimentally manipulated. Alternatively one could characterize naturally occurring environments with the help of these factors and study how people (should) act in them.

Moderator Cue to Expertise

The current study points to another important moderator variable in advice taking: the quality of the cue used to identify the better judge. As already pointed out by the PAR model, the relative advantage of choosing over averaging depends on the cue to expertise (Soll & Larrick, 2009). Particularly in the difficult task environment, the cue to expertise became crucial for the performance of choosing. Due to the lack of a good cue, averaging became better than choosing. Interesting enough, participants also adapted their preferred strategy in the direction of averaging. People thus seem to be sensitive to the naturally occurring (negative) intercorrelation between difficulty and the validity of confidence (which is not just the result of a smaller range of confidence ratings, see footnote 1; cf. Lichtenstein & Fischhoff, 1977). Thus not only is difficulty an intuitive way to change the environment but it might also signal to people the structure of the environment. In this regards, using representative advice was advantageous as it allowed us to study ecological rational behavior. On the other hand, the weak confidence cue in the difficult tasks did not allow us to study the actual performance of choosing in this environment. Future studies should investigate whether people resort more often to choosing if they believe or know they have a better cue to detect the better judge.

Because confidence plays a key role, interventions that aim at improving the use of advice could target at this cue. Providing people with a better cue would boost their performance because they would start choosing the better judge. Possible ways of improving the calibration of the judges' own confidence could be to provide judges with feedback about their own accuracy (whereas mere practice without feedback would not lead to a better calibration, cf. Paese & Sniezek, 1991). One way to improve the evaluation of an advisor's calibration would be to track performance together with confidence over some past experience and to communicate both to the judge (Soll & Mannes, 2011; see also Tenney, Spellman, & MacCoun, 2008). Also letting judge and advisor interact repeatedly may serve the same goal.

A less effortful alternative would be to more often bet on averaging. In fact, our results can be seen as further proof of the "robust beauty" of averaging (Dawes, 1979). Averaging has the advantage of not requiring any cues and of performing relatively well across all difficulty levels (if some bracketing is given; Soll & Larrick, 2009; Yaniv, 2004, study 1). Averaging might be an intuitive strategy when being confronted with difficult tasks, although otherwise being often misappreciated as a strategy for combining estimates (Larrick & Soll, 2006).

How Did People Decide How to Use Advice?

How did people manage to appropriately use choosing and averaging in the different task environments? In line with research on decision making (Gigerenzer et al., 1999), we assume that the environment provides cues which inform—if appropriately used—strategy choice. Which cues might have been used? One possible cue is the perceived difficulty. The perception of a task as being easy (for the majority of people), for example, may lead people to trust others' guesses and average them. Perceiving a task as difficult, in contrast, may cause people to hesitate to rely on others unless their expertise can be proved. Similarly perceived difficulty may inform about the trustworthiness of the cue confidence. Perceived difficulty may thus act as a meta-cue telling "kind" environments (when following the more confident is beneficial) apart from "wicked" environments (when confidence is rather misleading; cf. Hertwig, 2012; see also Koriat, 2012b).

A second starting point for deciding how to use advice is to take the distance between the own and the advisor's confidence into account. In fact, people were shown to be sensitive to differences in confidence: recommendations given by more confident advisors are followed more often than those given by their less confident advisors (Phillips, 1999; Sniezek & Buckley, 1995; Sniezek & Van Swol, 2001; Soll & Larrick, 2009; Van Swol & Sniezek, 2005; Yaniv, 1997). Moreover, some confidence values may be more informative than others; for example, the highest confidence value (in our study "6") may have a very persuasive effect because it is given rather rarely and if so, only when the judge is quite accurate (cf. Moussaïd, Kämmer, Pipergias Analytis, & Neth, 2012).

A different starting point could be to take the distance between the two estimates as a cue. Large distances may signal that it would pay more to identify the better judge. In case of having a (good) cue, judges may then tend to rather choose than average. Lacking a good cue as it would be the case in difficult tasks would trigger averaging. However, there might be an upper limit to the size of the distance between estimates in order to be considered at all: People were found to rather ignore advice with an extremely large distance (of, for example, 93 years in questions on historical data, Yaniv, 2004; which was termed the distance effect, Harries, Yaniv, & Harvey, 2004; Yaniv & Milyavsky, 2007).

Alternatively, people may use a combination of the distance between confidence ratings and between estimates. We (Moussaïd et al., 2012) showed in another set of studies, for example, that a decision tree describes people's advice-taking behavior well: In a first step, people evaluate the distance between estimates, and in a second step the difference between confidence ratings comes into play (see Supplementary Material of Moussaïd et al., 2012). For example, people combine estimates in case the distance between estimates is small and their own confidence is far below that of the advisor.

Benefits and Limitations

Besides being a systematic exploration of the impact of task difficulty on advice taking strategies, this study constitutes an application test of the PAR model by Soll and Larrick (2009). It proved to be a good theory for generating testable predictions about the relative performance of different strategies. By creating three different environments based on the perceived task difficulty, we successfully created natural environments with a wide variation in the relevant conditions identified by Soll and Larrick (2009), without resorting to simulated advice (Soll & Mannes, 2011).

A possible limitation of the current study (and of typical JAS) may be seen in the restricted availability of cues to expertise. In real life, people may have access to more and possibly more indicative cues, even for difficult tasks. For example, the advisor might also communicate the source of her knowledge, demonstrate the accuracy of her guess, or communicate the range in which the guess falls so that the judge has a better basis on which to evaluate the advice and can weight it accordingly (but see Yaniv & Foster, 1997 for how little people value such ranges). It might be an interesting research question of how people use multiple cues to expertise, especially when their relative validities change over different task environments. It could be that people employ a take-the-best strategy (Gigerenzer & Goldstein, 1996) when inferring the quality of advice before integrating it.

Limited generalizability of our results may be caused by the content of our experimental task. General knowledge questions may not be the most frequent situation in which we seek advice (if we are not in a game show). Most advice seeking probably happens in situations in which no true answer exists or is hardly accessible by anyone. We suppose that the ecological rationality approach offers a good framework for studying advice taking in such uncertain situations (cf. Todd et al., 2012).

Conclusion

What can be concluded is that future studies on advice taking should not focus only on the behavioral side, searching for explanations within the person but should also study the environmental side, as behavior is usually shaped by our resources, strategies, *and* the environment (cf. Simon, 1956). An insightful framework for this endeavor is Soll and Larrick’s (2009) PAR model. Taking an ecological rationality approach to study the impact of task difficulty on advice-taking behavior, we shed light on an important environmental factor that moderates the performance of advice taking strategies and shapes our behavior.

Acknowledgments

We are grateful to Gregor Caregnato and Jann Wäscher in collecting the data, and to Uwe Czienskowski for programming the software. We also thank Wolfgang Gaissmaier and Anselm Rothe for helpful discussions and Stephan Lewandowsky for his support and inspirations during the Summer School on Computational and Mathematical Modeling of Cognition, in 2012. Thanks to Gudrun Rauwolf and Tor Nielson for their support with creating the experimental task.

Chapter 5

General discussion

This dissertation explored how people make decisions under uncertainty together and with the help of others, and thus how the environment shapes their behavior. Core questions of the work presented here were (1) How do people make decisions within a social context? (2) Does social decision making differ from individual decision making? (3) What environmental factors influence the use and performance of strategies employed by groups for making decisions or by individuals for integrating advice? By taking an ecological rationality perspective to study these questions, in this dissertation I have contributed to both the group decision-making and the advice-taking literature and I have extended research on ecological rationality to different social contexts. In the following, I provide a summary of the contributions of the work reported in this dissertation.

Contributions: What Have We Learned from the Three Studies?

All three projects investigated adaptive strategy selection, each focusing on a different aspect of adaptivity. They thus contribute to the body of literature that provides evidence that individuals are able to select strategies that are appropriate for the environmental structure of a task (e.g., Bröder, 2003; Dieckmann & Rieskamp, 2007; Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006). In addition, the three projects extend these findings to more social contexts that so far have received little attention (for some exceptions see Reimer & Hoffrage, 2006; Reimer & Katsikopolous, 2004; Soll & Larrick, 2009).

Three core mechanisms have been discussed in the literature to explain how strategies are adaptively selected (cf. Marewski, 2010; Todd & Gigerenzer, 2012). The cognitive niche approach (Marewski & Schooler, 2011) suggests that constraints of memory capacity together with the available information determine the choice of heuristic. In all three projects, this principle lies at the heart of every strategy selection problem as it determines the initial set of heuristics to choose from. The cognitive niche approach is not, however, an explicit focus of this dissertation. A second explanation identifies reinforcement learning as a probable mechanism underlying adaptive strategy selection when a choice has to be made repeatedly from a set of heuristics (Rieskamp & Otto, 2006). The study reported in Chapter 2 is in the tradition of this approach and investigated how well and how fast individuals and dyads are able to learn to select the appropriate strategy when facing unfamiliar tasks with different underlying payoff structures and in changing environments. The third strategy selection mechanism, ecological rationality, focuses on how the structure of the environment influences the use and performance of heuristics. The adaptivity of the recognition heuristic is, for example, presumed to depend on the correlation of recognition with some environmental criterion, an issue I have elaborated on in Chapter 3. Furthermore, different environmental structures having an impact on the success of different advice-taking strategies, which was the explicit focus of the last study, reported in Chapter 4.

The studies, of course, cannot shed light on and do justice to all aspects of decision making in social contexts. Three main overall findings, however, can be derived from the research at hand.

The Environment Matters: Comparing Individuals and Dyads in Their Adaptive Use of Decision Strategies

In Chapter 2 I explored how well and how quickly groups were able to learn to select appropriate strategies in a given environment and to what extent groups differed from individuals. The core assumption of this study was that groups are information processors with cognition distributed across individuals (De Dreu, Nijstad, & van Knippenberg, 2008; Hinsz, Tindale, & Vollrath, 1997; Larson & Christensen, 1993).

In two experiments, two-member groups and individuals were confronted with an unfamiliar task. How quickly over repeated trials groups and individuals adapted their decision strategy was recorded. Adaptation was necessary because the environments differed in their payoff structure, making either take-the-best (Gigerenzer & Goldstein, 1999) or weighted additive (WADD; Dawes, 1979) the most successful strategy. Groups not only proved to be able to select the appropriate strategy, but—in the take-the-best-friendly environment—also showed slight superiority over individuals. Groups learned faster or reached a higher overall performance level. No differences, however, were found in the WADD-friendly environment where all groups and individuals performed very well and on average better than in the take-the-best environment.

The superiority of groups was expected because previous studies had already provided evidence for faster learning rates and a higher overall performance in intellective tasks (Hill, 1982; Hinsz et al., 1997). The advantage probably stems from a heightened ability to correct errors together with a larger memory capacity (Hinsz, 1990). The asymmetrical superiority, however, was surprising. What was it that made groups superior in the take-the-best-friendly environment but not in the WADD-friendly one? A number of possible reasons were discussed in Chapter 2, among others that the steps of information search and the rules of information integration might have been easier to verbalize in the take-the-best-friendly compared to the WADD-friendly environment. This would render take-the-best easier to communicate and teach to another person as soon as the need for this strategy had been detected by the best member.

In sum, this project drew the optimistic picture that individuals and groups can indeed learn to adapt their decision strategies to different environments if they have the opportunity to repeatedly encounter the task. More research, however, is needed to better understand the differences between individuals and groups in terms of their ability to learn adaptive strategies. It is, for example, necessary to explore additional domains and task characteristics, such as its familiarity, as well as the role of different group sizes and the way "real" groups make decisions in novel or changing environments.

The Adaptive Use of Recognition in Group Decision Making

Focusing on another facet of ecological rationality and connecting this concept to classic models of group decision making, the second study, reported in Chapter 3, explored the ecological rationality of group decision rules in terms of their fit to the composition of the group.

Heuristics are considered ecologically rational to the extent that they exploit the structure of the environment. The recognition heuristic, for example, exploits the correlation between recognition and some criterion and leads to correct predictions if that correlation is substantial (Goldstein & Gigerenzer, 2002). It seems that humans are sensitive to the size of correlations. When the correlation was high, for example, individuals were found to rely on recognition, but they discounted recognition when the correlation was low (e.g., Pachur & Hertwig, 2006; Pohl 2006; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011).

The question now arises, are small groups similarly sensitive to the validity of their members' recognition and knowledge and, as a consequence, able to adaptively select strategies that lead to the best achievable outcome by betting on either those members who use their knowledge or those who use their recognition? In an experiment, 43 three-member groups performed an inference task in which they had to infer which of two German companies had a higher market capitalization. From these choices, it could be concluded that, in the majority of trials, groups indeed applied the most successful strategy. Additional analyses of their discussions provided further information about the implementation of the strategies. For example, all groups highly valued the recognition heuristic as an argument. Moreover, groups that more frequently relied on their members using the recognition heuristic used the heuristic more often as an argument and more frequently at the very beginning of their discussions. As a side effect, the interactive experimental design allowed for studying how and how often the recognition heuristic was verbalized and used (cf. Hilbig, 2010).

In relation to this experiment, a number of possible mechanisms were identified that could underlie the adaptive use of the recognition heuristic. Common to all potential mechanisms is the assumption that humans have some access to the adaptivity of their own recognition validity (e.g., by their recognition fluency, cf.

Pachur, 2011; Pachur & Hertwig, 2006) which they then communicate to other members (e.g., by their confidence, knowledge cues, or by informing others about the source of their recognition). More research, however, is certainly needed to shed light on mechanisms of adaptive strategy selection in interacting groups. In addition, findings need to be extended from the rather restricted task type of paired comparisons to a much broader set of decision problems.

With the first two studies in Chapters 2 and 3, I concluded that the heuristics groups use are indeed similar to those individuals use and that groups perform well when they have to ignore less informative cues or less knowledgeable group members. That groups use simple heuristics to process information has been considered by only a few social psychologists (Gigone & Hastie, 1997). In contrast, the question of whether groups and individuals are susceptible to similar biases has received much more attention (for a review see Kerr, MacCoun, & Kramer, 1996). It seems to have been largely ignored that individuals and groups face very similar conditions in decision making, such as constraints in time and knowledge as well as uncertain environments, suggesting that heuristic decision making is a likely strategy to be used also by groups. Furthermore, the gold standard in much small-group research is the linear combination of all pieces of information, which are distributed across all members of a group (e.g., Stasser, Stewart, & Wittenbaum, 1995). The structure of the environment and the appropriateness of ignoring certain pieces of information, however, are rarely taken into consideration. More in-depth research could help to clarify what findings from individual decision making can be generalized to small groups and under what conditions group decision making proves ecologically rational. This is important not only on theoretical grounds, as “theory and research dealing with groups are relevant to nearly all the social sciences” (Forsyth, 1990, p.15), but also—and possibly more importantly—because of the great practical relevance of group work.

The Influence of Task Difficulty on Advice Taking: An Ecological Rationality Perspective

The third study, presented in Chapter 4, investigated the impact of task difficulty on the use of advice from an ecological rationality perspective. This third study resembled the second in that it also explored the impact of different environmental characteristics on behavior. Tasks of different difficulty levels were conceptualized as constituting different environments. The tasks proved to be related to certain predictable statistical properties; that is, easy tasks were found to be characterized by only small differences in accuracy between judges and a moderately high bracketing rate, while the opposite was true for difficult tasks.

The PAR model (Soll & Larrick, 2009) was applied to predict the accuracy of two prominent advice-taking strategies, namely, choosing and averaging, in the different task environments. Considering the two environmental factors identified by Soll and Larrick (2009), accuracy ratio and bracketing rate, I predicted that a strategy of averaging would perform well with easy tasks, while choosing the better judge would potentially outperform averaging in the case of difficult tasks. The likelihood of detecting the better judge was found to strongly moderate these relationships as a third factor. Because this factor was correlated with task difficulty, that is, because it became weaker with increasing difficulty, averaging turned out to be the most robust and best performing strategy not only for easy tasks, but also for difficult ones.

Furthermore, descriptive analyses revealed that people’s use of averaging differed depending on the level of task difficulty. As task difficulty increased, they resorted more and more to averaging (and to more mixed strategies) and thus appear to have been sensitive to their lack of a good cue to reliably detect who was the better judge.

This last study demonstrated the fruitfulness of adopting an ecological rationality perspective to study a topic that has received little attention in the advice-taking literature, even though it probably plays an important moderating role in almost every experiment, namely, task difficulty. By conceptualizing environments in terms of task difficulty, it was possible to derive testable hypotheses on the performance and use of strategies. As perceived task difficulty was “translated” into neutral statistical parameters (such as bracketing rate and accuracy ratio), it should be easy to further test our predictions in environments conceptualized in a way that differs from the one used here. Another interesting extension of the research at hand would be to study how advice from multiple advisors is used in different environments.

Gaps: What Do We Still Need to Explore?

While the studies presented in this dissertation provided valuable insights into the workings of how humans make adaptive decisions in groups and with the help of others, the findings indicate further investigations are needed. First and very generally, there is an obvious need for additional experimental studies to more comprehensively grasp both the boundary conditions of adaptive decision making in social contexts as well as the factors influencing those decisions. This could be done, for example, by more explicitly making use of the parallels between individual and group cognition, as was undertaken in the first study, in Chapter 2 (cf. Hinsz et al., 1997). In addition, a promising, indeed fundamental, research avenue to pursue is to closely link the concept of ecological rationality to models of group decision making. Study 2, in Chapter 3, has proposed a step in this direction (cf. Reimer & Katsikopoulos, 2004).

Second and more specifically, the connections between formal concepts of simple heuristics (e.g., Katsikopoulos & Martignon, 2006) and of group decision rules and their applicability (e.g., Einhorn, Hogarth, & Klemperer, 1978; Hastie & Kameda, 2005) have to be worked out more precisely. This would not only strengthen the theoretical basis for further experiments but would also allow for the derivation of testable hypotheses inspired by the respective other discipline. One promising example is the extrapolation from the ecological rationality of individual decision making to the "ecological rationality of social combination rules" (Reimer & Hoffrage, 2012, p. 356). Moreover, and as the three studies showed, it would be productive in this respect to apply more formal model testing to the field of social psychology.

Third and with immediate practical implications, it is crucial to examine and understand more about the way humans search for and integrate information "in the wild." One such real-world case of great practical, indeed societal relevance can be found in the medical domain, namely, in diagnostic decision making. An extension to this domain of the research presented in this dissertation is planned for next year and will cover such research questions as, Does the quality of diagnoses increase if they are made by a small group instead of a single person? Does the selection of diagnostic tests change if a small group instead of a single person is involved? To what extent and in what way do hierarchical differences (e.g., two physicians of different status working together) in the level of experience or accountability affect the decision-making process leading to a diagnosis?

A nuanced understanding of how people make decisions together with others is crucial to facilitate collaborations and to make wise use of the potential groups that individuals can form. This dissertation provided some examples of how applying the framework of ecological rationality can contribute to this endeavor.

References

References

Abele, S., Stasser,S. & Groebe, M. (2012). When expertise trumps extraversion: The role of functional diagnosticity. *Seventh Annual INGroup Conference, Chicago, Illinois, U.S.A*.

Allen, N. J., & Hecht, T. D. (2004). The 'romance of teams': Toward an understanding of its psychological underpinnings and implications. *Journal of Occupational & Organizational Psychology, 77*, 439-461. doi: 10.1348/0963179042596469

Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology, 96*, 491-503. doi: 10.1037/a0014201

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 19 (pp. 1-40). New York: Academic Press.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396-403. doi: 10.1111/j.1745-6916.2007.00051.x

Bell, S. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology, 92*, 595-615. doi: 10.1037/0021-9010.92.3.595

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 107-129. doi: 10.1037/0278-7393.33.1.107

Betsch, T., Fiedler, K., & Brinkmann, J. (1998). Behavioral routines in decision making: The effects of novelty in task presentation and time pressure on routine maintenance and deviation. *European Journal of Social Psychology, 28*, 861-878. doi: 10.1002/(SICI)1099-0992(1998110)28:6<861::AID-EJSP899>3.0.CO;2-D

Betsch, T., & Haberstroh, S. (2005). *The routines of decision making*. Mahwah, NJ: Erlbaum.

Betsch, T., Haberstroh, S., Glöckner, A., Haar, T., & Fiedler, K. (2001). The effects of routine strength on adaptation and information search in recurrent decision making. *Organizational Behavior and Human Decision Processes, 84*, 23-53. doi: 10.1006/obhd.2000.2916

Betsch, T., Haberstroh, S., & Höhle, C. (2002). Explaining routinized decision making. *Theory & Psychology, 12*, 453-488. doi: 10.1177/0959354302012004294

Betts, K. R., & Hinsz, V. B. (2010). Collaborative group memory: Processes, performance, and techniques for improvement. *Social and Personality Psychology Compass, 4*, 119-130. doi: 10.1111/j.1751-9004.2009.00252.x

Biele, G., & Rieskamp, J. (in press). The advice of others: When and how we benefit from it. In R. Hertwig, U. Hoffrage, & the ABC Research Group (Eds.), *Simple heuristics in a social world*. New York: Oxford University Press.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*, 127-151. doi: 10.1016/j.obhdp.2006.07.001

Bonner, B. L. (2004). Expertise in group problem solving: Recognition, social combination, and performance. *Group Dynamics: Theory, Research, and Practice, 8*, 277-290. doi: 10.1037/1089-2699.8.4.277

Bottger, P., & Yetton, P. (1988). An integration of process and decision scheme explanations of group problem solving performance. *Organizational behavior and human decision processes, 42*, 234-249. doi: 10.1016/0749-5978(88)90014-3

Bowers, J. S., & Davis, C. J. (2012a). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin, 138,* 389–414. doi:10.1037/a0026450

Bowers, J. S., & Davis, C. J. (2012b). Is that what Bayesians believe? Reply to Griffiths, Chater, Norris, and Pouget (2012). *Psychological Bulletin, 138*, 423–426. doi:10.1037/a0027750

Bröder, A. (2003). Decision making with the 'adaptive toolbox': Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 611-625. doi: 10.1037/0278-7393.29.4.611

Bröder, A. & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making, 16*, 193-213. doi: 10.1002/bdm.442

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 904-918. doi: 10.1037/0278-7393.32.4.904

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91*, 1189-1207. doi: 10.1037/0021-9010.91.6.1189

Chalos, P., & Pickard, S. (1985). Information choice and cue use: An experiment in group information-processing. *Journal of Applied Psychology, 70*, 634-641. doi: 10.1037/0021-9010.70.4.634

Christensen-Szalanski, J. J. (1978). Problem solving strategies: A selection mechanism, some implications, and some data. *Organizational Behavior & Human Performance*, *22,* 307-323. doi: 10.1016/0030-5073(78)90019-3

Christensen-Szalanski, J. J. (1980). A further examination of the selection of problem-solving strategies: The effects of deadlines and analytic aptitudes. *Organizational Behavior & Human Performance, 25*, 107-122. doi: 10.1016/0030-5073(80)90028-8

Chu, P., & Spires, E. E. (2003). Perceptions of accuracy and effort of decision strategies. *Organizational Behavior and Human Decision Processes, 91*, 203-214. doi: 10.1016/S0749-5978(03)00056-6

Cooper, R. (2000). Simple heuristics could make us smart; but which heuristics do we apply when? *Behavioral and Brain Sciences, 23*, 746.

Creyer, E. H., Bettman, J. R., & Payne, J. W. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making, 3*, 1-16. doi: 10.1002/bdm.3960030102

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97-118). New York: Oxford University Press.

Czienskowski, U. (2004). The oil drilling experiment [Computer software]. Berlin, Germany: Max Planck Institute for Human Development.

Davis, J. H. (1969). Individual-group problem solving, subject preference, and problem type. *Journal of Personality and Social Psychology, 13*, 362-374. doi: 10.1037/h0028378

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review, 80*, 97-125. doi: 10.1037/h0033951

Davis, J. H. (1992). Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples, 1950-1990. *Organizational Behavior and Human Decision Processes, 52*, 3-38. doi: 10.1016/0749-5978(92)90044-8

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist, 34*, 571-582. doi: 10.1037/0003-066X.34.7.571

De Dreu, C. K. W., Nijstad, B. A., & van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review, 12*, 22-49. doi: 10.1177/1088868307304092

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dieckmann, A., & Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition, 35*, 1801-1813. doi: 10.3758/BF03193511

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology, 53*, 497-509. doi: 10.1037/0022-3514.53.3.497

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171-192. doi: 10.1016/0030-5073(75)90044-6

Erdfelder, E., Küpper-Tetzel, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making, 6,* 7-22.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314. doi: 10.1016/j.tics.2004.05.002

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications Ltd.

Forsyth, D. R. (1990). *Group dynamics* (2nd ed.). Belmont: Brooks & Cole.

Frey, D., Schulz-Hardt, S., & Stahlberg, D. (1996). Information seeking among individuals and groups and possible consequences for decision making in business and politics *Understanding group behavior, Vol 2: Small group processes and interpersonal relations* (pp. 211-225). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; England.

Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. *Judgment and Decision Making, 6*, 73-88.

Galton, F. (1907). Vox populi. *Nature, 75*, 450-451. doi: 10.1038/075450a0

Garcia-Retamero, R., Takezawa, M., & Gigerenzer, G. (2009). Does imitation benefit cue order learning? *Experimental Psychology, 56*, 307-320. doi: 10.1027/1618-3169.56.5.307

Gersick, C. J. G., & Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes, 47*, 65-97. doi: 10.1016/0749-5978(90)90047-D

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster. (UK edition: *Reckoning with risk. Learning to live with uncertainty*. London: Penguin.)

Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 62-88). Malden: Blackwell.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482. doi: 10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669. doi:

10.1037/0033-295X.103.4.650

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 75-96). New York, NY: Oxford University Press.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making, 6*, 100-121.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Gigone, D., & Hastie, R. (1997). The impact of information on small group choice. *Journal of Personality and Social Psychology, 72*, 132-140. doi: 10.1037/0022-3514.72.1.132

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology, 102*, 497-512. doi: 10.1037/a0026413

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making, 20*, 21-35. doi: 10.1002/bdm.539

Glöckner, A., & Betsch, T. (2008). Multipe-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1055-1075. doi: 10.1037/0278-7393.34.5.1055

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making, 23*, 439-462. doi: 10.1002/bdm.668

Glossary of Deutsche Börse Group. (n.d.). Market capitalization. Retrieved February 20, 2008, from http://deutsche-boerse.com/dbag/dispatch/en/kir/gdb_navigation/info_center/40_Know_how/10_Stock_Exchange_A_Z?glossaryWord=pi_glos_marktkapitalisierung

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*, 75-90. doi: 10.1037/0033-295X.109.1.75

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin, 138*, 415–422. doi:10.1037/a0026884

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology, 47*, 307-338. doi: 10.1146/annurev.psych.47.1.307

Hackman, J. R., Brousseau, K. R., & Weiss, J. A. (1976). The interaction of task design and group performance strategies in determining group effectiveness. *Organizational Behavior & Human Performance, 16*, 350-365. doi: 10.1016/0030-5073(76)90021-0

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: The weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making, 17*, 333-348. doi: 10.1002/bdm.474

Harvey, N., & Bolger, F. (2001). Collecting information: Optimizing outcomes, screening options, or facilitating discrimination? *Quarterly Journal of Experimental Psychology: Section A, 54*, 269-301. doi: 10.1080/02724980042000110

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes, 70*, 117-133. doi: 10.1006/obhd.1997.2697

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Decision research* (pp. 129-157). Greenwich, CT: JAI Press.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review, 112*, 494-508. doi: 10.1037/0033-295X.112.2.494

Henry, R. A. (1995). Improving group judgment accuracy: Information sharing and determining the best member. *Organizational Behavior and Human Decision Processes, 62*, 190-197. doi: 10.1006/obhd.1995.1042

Henry, R. A., Strickland, O. J., Yorges, S. L., & Ladd, D. (1996). Helping groups determine their most accurate member: The role of outcome feedback. *Journal of Applied Social Psychology, 26*(13), 1153-1170. doi: 10.1111/j.1559-1816.1996.tb02290.x

Hertwig, R. (2012) Tapping into the wisdom of the crowd - with confidence. *Science, 336*, 303-304. doi: 10.1126/science.1221403

Hertwig, R., & Herzog, S. M. (2009). Fast and frugal heuristics: Tools of social rationality. *Social Cognition, 27*, 661-698. doi: 10.1521/soco.2009.27.5.661

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1191-1206. doi: 10.1037/a0013025

Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation. In: G. Gigerenzer, P. Todd, and the ABC Research Group (Eds.). *Simple heuristics that make us smart* (pp. 209-234). Oxford University Press, New York.

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 213-231). Chichester, England: Wiley. doi: 10.1002/047001332X.ch11

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*, 231-237. doi: 10.1111/j.1467-9280.2009.02271.x

Hilbig, B. E. (2010). Reconsidering 'evidence' for fast and frugal heuristics. *Psychonomic Bulletin & Review, 17*, 923-930. doi: 10.3758/PBR.17.6.923

Hilbig, B. E., & Pohl, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology, 55*, 394-401. doi: 10.1027/1618-3169.55.6.394

Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol.35*, 1296-1305. doi: 10.1037/a0016565

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91*, 517-539. doi: 10.1037/0033-2909.91.3.517

Hinsz, V. B. (1990). Cognitive and consensus processes in group recognition memory performance. *Journal of Personality and Social Psychology, 59*, 705-718. doi: 10.1037/0022-3514.59.4.705

Hinsz, V. B., & Davis, J. H. (1984). Persuasive arguments theory, group polarization, and choice shifts. *Personality and Social Psychology Bulletin, 10*, 260-268.

doi: 10.1177/0146167284102012

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*, 43-64. doi: 10.1037/0033-2909.121.1.43

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance, 21*, 40-46. doi: 10.1016/0030-5073(78)90037-5

Hogarth, R. M., & Karelaia, N. (2006). "Take-the-best" and other simple strategies: Why and when they work "well" with binary cues. *Theory and Decision, 61*, 205-249. doi: 10.1007/s11238-006-9000-8

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review; Psychological Review, 114*, 733-758. doi: 10.1037/0033-295X.114.3.733

Huber, O., & Seiser, G. (2001). Accounting and convincing: The effect of two types of justification on the decision process. *Journal of Behavioral Decision Making, 14*, 69-85. doi: 10.1002/1099-0771(200101)14:1<69::AID-BDM366>3.0.CO;2-T

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34*, 169–231. doi:10.1017/S0140525X10003134

Kämmer, J. E., Gaissmaier, W., & Czienskowski, U. (2012). The environment matters: Comparing individuals and dyads in their adaptive use of decision strategies. *Manuscript submitted for publication*.

Kämmer, J. E., Gaissmaier, W., Reimer, T., & Schermuly, C. C. (2012). The adaptive use of recognition in group decision making. *Manuscript submitted for publication*.

Kämmer, J. E., Pipergias Analytis, P., Neth, H., & Moussaid, M. (2012). Social influence and the moderating effect of task difficulty on the wisdom of crowds. *Poster presented at the 54th meeting of The Experimental Approach to Psychology (TeaP)*, Mannheim, Germany.

Karau, S. J., & Kelly, J. R. (1992). The effects of time scarcity and time abundance on group performance quality and interaction process. *Journal of Experimental Social Psychology, 28*, 542-571. doi: 10.1016/0022-1031(92)90045-L

Karau, S. J. & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*, 681-706. doi: 10.1037/0022-3514.65.4.681

Keller, M., Gummerum, M., Canz, T., Gigerenzer, G., & Takezawa, M. (in press). The “is” and “ought” of sharing: The equality heuristic across the lifespan. In R. Hertwig, U. Hoffrage, & the ABC Research Group, *Simple heuristics in the social world*. New York, NY: Oxford University Press.

Kelly, J. R., & Loving, T. J. (2004). Time pressure and group performance: Exploring underlying processes in the attentional focus model. *Journal of Experimental Social Psychology, 40*, 185-198. doi: 10.1016/S0022-1031(03)00094-5

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103*, 687-719. doi: 10.1037/0033-295X.103.4.687

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*, 623-655. doi: 10.1146/annurev.psych.55.090902.142009

Kocher, M. G., & Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal, 115*, 200-223. doi: 10.1111/j.1468-0297.2004.00966.x

Koriat, A. (2012a). The self-consistency model of subjective confidence. *Psychological Review, 119*, 80-113. doi: 10.1037/a0025648

Koriat, A. (2012b). When are two heads better than one and why? *Science, 336*, 360-362. doi: 10.1126/science.1216549

Krause, S., James, R., Faria, J. J., Ruxton, G. D., & Krause, J. (2011). Swarm intelligence in humans: diversity can trump ability. *Animal Behaviour, 81*, 941-948. doi: 10.1016/j.anbehav.2010.12.018

Kruger, J. (1999). Lake Wobegon be gone! The" below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221-232. doi: 10.1037/0022-3514.77.2.221

Ladbury, J. L., & Hinsz, V. B. (2009). Individual expectations for group decision processes: Evidence for overestimation of majority influence. *Group*

Dynamics: Theory, Research, and Practice, 13, 235-254. doi: 10.1037/a0015424

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*, 111-127. doi. 10.1287/mnsc.1050.0459

Larson, J. R., Foster-Fishman, P. G., & Keys, C. B. (1994). Discussion of shared and unshared information in decision-making groups. *Journal of Personality and Social Psychology, 67*, 446-461. doi: 10.1037/0022-3514.67.3.446

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology, 37*, 822-832. doi: 10.1037/0022-3514.37.6.822

Laughlin, P. R. (1980). Social combination processes of cooperative problem-solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 127-155). Hillsdale, NJ: Earlbaum.

Laughlin, P. R. (1999). Collective induction: Twelve postulates. *Organizational Behavior and Human Decision Processes*, *80*, 50-69.

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88*, 605-620. doi: 10.1016/S0749-5978(02)00003-1

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22*, 177-189. doi: 10.1016/0022-1031(86)90022-3

Laughlin, P. R., & Shippy, T. A. (1983). Collective induction. *Journal of Personality and Social Psychology, 45*, 94-100. doi: 10.1037/0022-3514.45.1.94

Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology, 61*, 50-67. doi: 10.1037/0022-3514.61.1.50

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review, 11*, 343-352.

LePine, J. A. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology, 88*, 27-39. doi: 10.1037/0021-9010.88.1.27

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*, 255-275. doi: 10.1037/0033-2909.125.2.255

Levine, J. M., & Smith, E. R. (in press). Group cognition: Collective information search and distribution. In D. Carlston (Ed.), *Oxford handbook of social cognition*. New York: Oxford University Press.

Libby, R., Trotman, K. T., & Zimmer, I. (1987). Member variation, recognition of expertise, and group performance. *Journal of Applied Psychology, 72*, 81-87. doi: 10.1037/0021-9010.72.1.81

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159-183. doi: 10.1016/0030-5073(77)90001-0

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences, 108*(22), 9020-9025. doi: 10.1073/pnas.1008636108

Lorge, I., & Solomon, H. (1955). Two models of group behavior in the solution of eureka-type problems. *Psychometrika, 20*, 139-148.

Maciejovsky, B., Sutter, M., Budescu, D. V., & Bernau, P. (2010). Teams make you smarter: Learning and knowledge transfer in auctions and markets by teams and individuals. IZA Discussion Paper No. 5105. Available at SSRN: http://ssrn.com/abstract=1659084

Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: A review of the literature. *Acta Anaesthesiologica Scandinavica, 53*, 143-151. doi: 10.1111/j.1399-6576.2008.01717.x

Marewski, J. N. (2010). On the theoretical precision and strategy selection problem of a single-strategy approach: A comment on Glöckner, Betsch, and Schindler (2010). *Journal of Behavioral Decision Making, 23*, 463-467. doi: 10.1002/bdm.680

Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2009). Do voters use episodic knowledge to rely on recognition? In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference* *of the Cognitive Science Society* (pp. 2232–2237). Austin, TX: Cognitive Science Society.

Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multi-alternative inference. *Psychonomic Bulletin and Review, 17,* 287–309.

Marewski, J. N., Pohl, R. F., & Vitouch, O. (Eds.). (2010). Recognition processes in inferential decision making [Special issue]. *Judgment and Decision Making, 5*(4).

Marewski, J. N., Pohl, R. F., & Vitouch, O. (Eds.). (2011a). Recognition processes in inferential decision making (II) [Special issue]. *Judgment and Decision Making, 6*(1).

Marewski, J. N., Pohl, R. F., & Vitouch, O. (Eds.). (2011b). Recognition processes in inferential decision making (III) [Special issue]. *Judgment and Decision Making, 6*(5).

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review, 118*, 393-437. doi: 10.1037/a0024143

Marewski J. N., Schooler L. J. & Gigerenzer G. (2010). Five principles for studying people's use of heuristics. *Acta Psychologica Sinica*, *42*, 72-87.

McAllister, D. W., Mitchell, T. R., & Beach, L. R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance, 24*, 228-244. doi: 10.1016/0030-5073(79)90027-8

Meyer, B., Shemla, M., & Schermuly, C. C. (2011). Social category salience moderates the effect of diversity faultlines on information elaboration. *Small Group Research*, *42*, 257-282. doi: 10.1177/1046496411398396

Miner, F. C. (1984). Group versus individual decision making: An investigation of performance measures, decision strategies, and process losses / gains. *Organizational Behavior and Human Performance, 33*, 112-124. doi: 10.1016/0030-5073(84)90014-X

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85*, 1121-1135. doi:

10.1037/0022-3514.85.6.1121

Moussaid, M., Kämmer, J. E., Pipergias Analytis, P., & Neth, H. (2012). Social influence and the collective dynamics of opinion formation. *Manuscript submitted for publication.*

Müller-Trede, J. (2011). Repeated judgment sampling: Boundaries. *Judgment and Decision Making, 6*, 283-294.

Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance, 23*, 309-338. doi: 10.1016/0030-5073(79)90001-1

Newell, B. R., & Lee, M. D. (2011). The right tool for the job? Comparing an evidence accumulation and a naive strategy selection model of decision making. *Journal of Behavioral Decision Making, 24*, 456-481. doi: 10.1002/bdm.703

Newell, B. R., & Shanks, D. R. (2003). Take the best or at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 53-65. doi: 10.1037/0278-7393.29.1.53

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes, 91*, 82-96. doi: 10.1016/S0749-5978(02)00525-3

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259. doi: 10.1037/0033-295X.84.3.231

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*, 1-18. doi: 10.1016/0022-2496(66)90002-2

Olsson, A. C., Juslin, P., & Olsson, H. (2006). Individuals and dyads in a multiple-cue judgment task: Cognitive processes and performance. *Journal of Experimental Social Psychology, 42*, 40-56. doi: 10.1016/j.jesp.2005.01.004

Pachur, T. (2011). The limited value of precise tests of the recognition heuristic. *Judgment and Decision Making, 6*, 413-422.

Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making, 21*, 183-210. doi: 10.1002/bdm.581

Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 983-1002. doi: 10.1037/0278-7393.32.5.983

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Psychology, 2*, 147. doi: 10.3389/fpsyg.2011.00147

Paese, P. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational Behavior and Human Decision Processes, 48*, 100-130. doi: 10.1016/0749-5978(91)90008-H

Page, S. E. (2007). Making the difference: Applying a logic of diversity. *The Academy of Management Perspectives, 21*, 6-20. doi: 10.5465/AMP.2007.27895335

Parks, C. D., & Cowlin, R. A. (1996). Acceptance of uncommon information into group discussion when that information is or is not demonstrable. *Organizational Behavior and Human Decision Processes, 66*, 307-315. doi: 10.1006/obhd.1996.0058

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534-552. doi: 10.1037/0278-7393.14.3.534

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, England: Cambridge University Press.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 85-92. doi: 10.1037/0278-7393.14.1.85

Peterson, R.S., Owens, P.D., Tetlock, P.E., Fan, E.T., & Martorana, P. (1998). Group dynamics in top management teams: Groupthink, vigilance, and alternative models of organizational failure and success. *Organizational Behavior and Human Decision Processes, 73*, 272–305. doi: 10.1006/obhd.1998.2763

Phillips, J. M. (1999). Antecedents of leader utilization of staff input in decision-making teams. *Organizational Behavior and Human Decision Processes, 77*, 215-242. doi: 10.1006/obhd.1998.2819

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19*, 251-271. doi: 10.1002/bdm.522

Pohl, R. F. (2011). On the use of recognition in inferential decision making: An overview of the debate. *Judgment and Decision Making, 6*, 423-438.

Randall, K. R., Resick, C. J., & DeChurch, L. A. (2011). Building team adaptive capacity: The roles of sensegiving and team composition. *Journal of Applied Psychology, 96*, 525-540. doi: 10.1037/a0022622

Reimer, T., Bornstein, A.-L., & Opwis, K. (2005). Positive and negative transfer effects in groups. In T. Betsch & S. Haberstroh (Eds.), *The routine of decision making* (pp. 175-192). Mahwah, NJ: Erlbaum.

Reimer, T., Hertwig, R., & Sipek, S. (in press). Probabilistic persuasion: A Brunswikian theory of argumentation. In R. Hertwig, U. Hoffrage, & the ABC Research Group, *Simple heuristics in a social world*. New York, NY: Oxford University Press.

Reimer, T., & Hoffrage, U. (2003). *Information aggregation in groups: The approach of simple group heuristics (SIGH)*. In: R. Alterman, D. Kirsch (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (pp. 982–987). Cognitive Science Society, Boston.

Reimer, T., & Hoffrage, U. (2006). The ecological rationality of simple group heuristics: Effects of group member strategies on decision accuracy. *Theory and Decision, 60*, 403-438. doi: 10.1007/s11238-005-4750-2

Reimer, T., & Hoffrage, U. (2012). Ecological rationality for teams and committees: Heuristics in group decision making. In P. M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world* (pp. 335-359). New York: Oxford University Press.

Reimer, T., Hoffrage, U., & Katsikopoulos, K. V. (2007). Entscheidungsheuristiken in Gruppen [Heuristics in group decision-making]. *NeuroPsychoEconomics, 2*, 7-29.

Reimer, T., & Katsikopoulos, K. V. (2004). The use of recognition in group decision-making. *Cognitive Science, 28*, 1009-1029. doi: 10.1016/j.cogsci.2004.06.004

Richter, A., Dawson, J., & West, M. (2011). The effectiveness of teams in organizations: A meta-analysis. *The International Journal of Human Resource Management, 22*, 2749-2769. doi: 10.1080/09585192.2011.573971

Rieskamp, J., & Dieckmann, A. (2012). Redundancy: Environment structure that simple heuristics can exploit. In P. M. Todd, G. Gigerenzer & t. A. R. Group (Eds.), *Ecological rationality: Intelligence in the world* (pp. 187-215). New York: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart* (pp. 141-167). New York, NY: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica, 127*, 258-276. doi: 10.1016/j.actpsy.2007.05.004

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135*, 207-236. doi: 10.1037/0096-3445.135.2.207

Salas, E., Cooke, N., & Rosen, M. (2008). On teams, teamwork, and team performance: Discoveries and developments. Human Factors, 50, 540-547. doi:10.1518/001872008X288457

Savadori, L., Van Swol, L. M., & Sniezek, J. A. (2001). Information sampling and confidence within groups and judge advisor systems. *Communication Research, 28*, 737-771. doi: 10.1177/009365001028006002

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (in press). Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological Review*.

Schermuly, C. C., & Scholl, W. (2011). *IKD—Instrument zur Kodierung von Diskussionen [DCS—Discussion Coding System]*. Göttingen, Germany: Hogrefe.

Schermuly, C. C., & Scholl, W. (2012). The Discussion Coding System (DCS). A new instrument for analyzing communication processes. *Communication Methods and Measures, 6,* 12-40. doi: 10.1080/19312458.2011.651346

Scholl, W. (2005). Grundprobleme der Teamarbeit und ihre Bewältigung - Ein Kausalmodell [Basic problems of team work and their solution - a causal model]. In M. Högl & H. G. Gemünden (Eds.), *Management von Teams.*

Theoretische Konzepte und empirische Befunde (pp. 33-66). Wiesbaden: Gabler.

Scholl, W. (2007). Plädoyer für eine sozialere und dadurch interdisziplinärere und anwendbarere Sozialpsychologie [Plea for a more social, interdisciplinary, and applicable social psychology]. *Zeitschrift für Sozialpsychologie, 38*, 285-296. doi: 10.1024/0044-3514.38.4.285

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*, 610-628. doi: 10.1037/0033-295X.112.3.610

Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets.* Master of Science, University of Pittsburgh, Pittsburgh.

Shaw, M. E. (1961). A serial position effect in social influence on group decisions. *The Journal of Social Psychology, 54*, 83-91. doi: 10.1080/00224545.1961.9919353

Shiflett, S. (1972). Group performance as a function of task difficulty and organizational interdependence. *Organizational Behavior and Human Performance, 7*, 442-456. doi: 10.1016/0030-5073(72)90028-1

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138. doi: 10.1037/h0042769

Sjöberg, L. (2009). Are all crowds equally wise? A comparison of political election forecasts by experts and the public. *Journal of Forecasting, 28*, 1-18. doi: 10.1002/for.1083

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes, 62*, 159-174. doi: 10.1006/obhd.2000.2926

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes, 43*, 1-28. doi: 10.1016/0749-5978(89)90055-1

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes, 84*, 288-307. doi: 10.1006/obhd.2000.2926

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology, 38*, 317-346. doi: 10.1006/cogp.1998.0699

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 780. doi: 10.1037/a0015145

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting, 27*, 81-102. doi: 10.1016/j.ijforecast.2010.05.003

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222. doi: 10.1080/14640747308400340

Stasser, G. (1992). Information salience and the discovery of hidden profiles by decision-making groups: A 'thought experiment.'. *Organizational Behavior and Human Decision Processes, 52*, 156-181. doi: 10.1016/0749-5978(92)90049-D

Stasser, G. (2012). Missing links in small group decisions: Participation and recognition of expertise. *Invited Talk at the Max Planck Institute for Human Development, Adaptive Behavior and Cognition. Berlin, Germany*.

Stasser, G., & Davis, J. H. (1981). Group decision making and social influence: A social interaction sequence model. *Psychological Review, 88*, 523-551. doi: 10.1037/0033-295X.88.6.523

Stasser, G., Taylor, L. A., & Hanna, C. (1989). Information sampling in structured and unstructured discussions of three- and six-person groups. *Journal of Personality and Social Psychology, 57*, 67-78. doi: http://dx.doi.org/10.1037/0022-3514.57.1.67

Stasser, G., & Titus, W. (2003). Hidden profiles: A brief history. *Psychological Inquiry, 14*, 304-313. doi: 10.1080/1047840X.2003.9682897

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.

Stone, T. (1971). Effects of mode of organization and feedback level on creative task groups. *Journal of Applied Psychology, 55*, 324-330. doi: 10.1037/h0031534

Stroebe, W., Nijstad, B. A., & Rietzschel, E. F. (2010). Beyond productivity loss in brainstorming groups: The evolution of a question. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (pp. 157-203). San Diego: Elsevier.

Surowiecki, J. (2004). The wisdom of crowds. Why the many are smarter than the few and how collective wisdom shapes business, economics, societies, and nations.

New York: Doubleday.

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychologica, 80*, 143-168. doi: 10.1016/0001-6918(92)90044-E

Svenson, O., Edland, A., & Slovic, P. (1990). Choices and judgments of incompletely described decision alternatives under time pressure. *Acta Psychologica, 75*, 153-169. doi: 10.1016/0001-6918(90)90084-S

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology, 44*, 1368-1375. doi: 10.1016/j.jesp.2008.04.006

Tindale, R. S. (1989). Group vs. individual information processing: The effects of outcome feedback on decision making. *Organizational Behavior and Human Decision Processes, 44*, 454-473. doi: 10.1016/0749-5978(89)90019-8

Tindale, R. S., & Sheffey, S. (2002). Shared information, cognitive load, and group memory. *Group Processes & Intergroup Relations, 5*, 5-18. doi: 10.1177/1368430202005001535

Todd, P. M., & Gigerenzer, G. (2012). What is ecological rationality? In P. M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world* (pp. 3-30). New York: Oxford University Press.

Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: Intelligence in the world*. New York: Oxford University Press.

Tukey, J.W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology, 44*, 443-461. doi: 10.1348/014466604X17092

Vollrath, D. A., Sheppard, B. H., Hinsz, V. B., & Davis, J. H. (1989). Memory performance by decision-making groups and individuals. *Organizational Behavior and Human Decision Processes, 43*, 289-300. doi: 10.1016/0749-5978(89)90040-X

Volz, K. G., Schooler, L. J., Schubotz, R. I., Raab, M., Gigerenzer, G., & von Cramon, D. (2006). Why you think Milan is larger than Modena: Neural correlates of the recognition heuristic. *Journal of Cognitive Neuroscience, Vol.18*, 1924-

1936. doi: 10.1162/jocn.2006.18.11.1924

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science, 19*, 645-647. doi: 10.1111/j.1467-9280.2008.02136.x

Waller, M. J. (1999). The timing of adaptive group responses to nonroutine events. *Academy of Management Journal, 42*, 127-137.

Watson, G. B. (1928). Do groups think more efficiently than individuals? *The Journal of Abnormal and Social Psychology, 23*(3), 328-336. doi: 10.1037/h0072661

Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2009). Smart strategies for doctors and doctors-in-training: Heuristics in medicine. *Medical Education, 43,* 721–728.

Weldon, M. S., Blair, C., & Huebsch, P. D. (2000). Group remembering: Does social loafing underlie collaborative inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1568-1577. doi: 10.1037/0278-7393.26.6.1568

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The influence of egocentrism and focalism on people's optimism in competitions: when what affects us equally affects me more. *Journal of Personality and Social Psychology, 85*, 389-408. doi: 10.1037/0022-3514.85.3.389

Winquist, J. R., & Larson, J. R. J. (1998). Information pooling: When it impacts group decision making. *Journal of Personality and Social Psychology, 74*, 371-377. doi: 10.1037/0022-3514.74.2.371

Wittenbaum, G. M., & Stasser, G. (1996). Management of information in small groups. In J. L. Nye, A. M. Brower, J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition? Research on socially shared cognition in small groups.* (pp. 3-28). Thousand Oaks, CA US: Sage Publications, Inc.

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes, 69*, 237-249. doi: 10.1006/obhd.1997.2685

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*, 1-13. doi: 10.1016/j.obhdp.2003.08.002

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation.

Journal of Behavioral Decision Making, 10, 21-32. doi: 0894-3257/97/010021-12

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83*, 260-281. doi: 10.1006/obhd.2000.2909

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes, 103*, 104-120. doi: 10.1016/j.obhdp.2006.05.006

Yetton, P. W., & Bottger, P. C. (1982). Individual versus group problem solving: An empirical test of a best-member strategy. *Organizational Behavior and Human Performance, 29*, 307-321. doi: 10.1016/0030-5073(82)90248-3

Zakay, D. (1985). Post-decisional confidence and conflict experienced in a choice process. *Acta Psychologica, 58*, 75-80. doi: 10.1016/0001-6918(85)90035-6

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98*, 103-115. doi: 10.1016/S0001-6918(97)00037-1

Zarnoth, P., & Szniezek, J. A. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology, 33*, 345-366. doi: 10.1006/jesp.1997.1326

Appendices

Appendix A

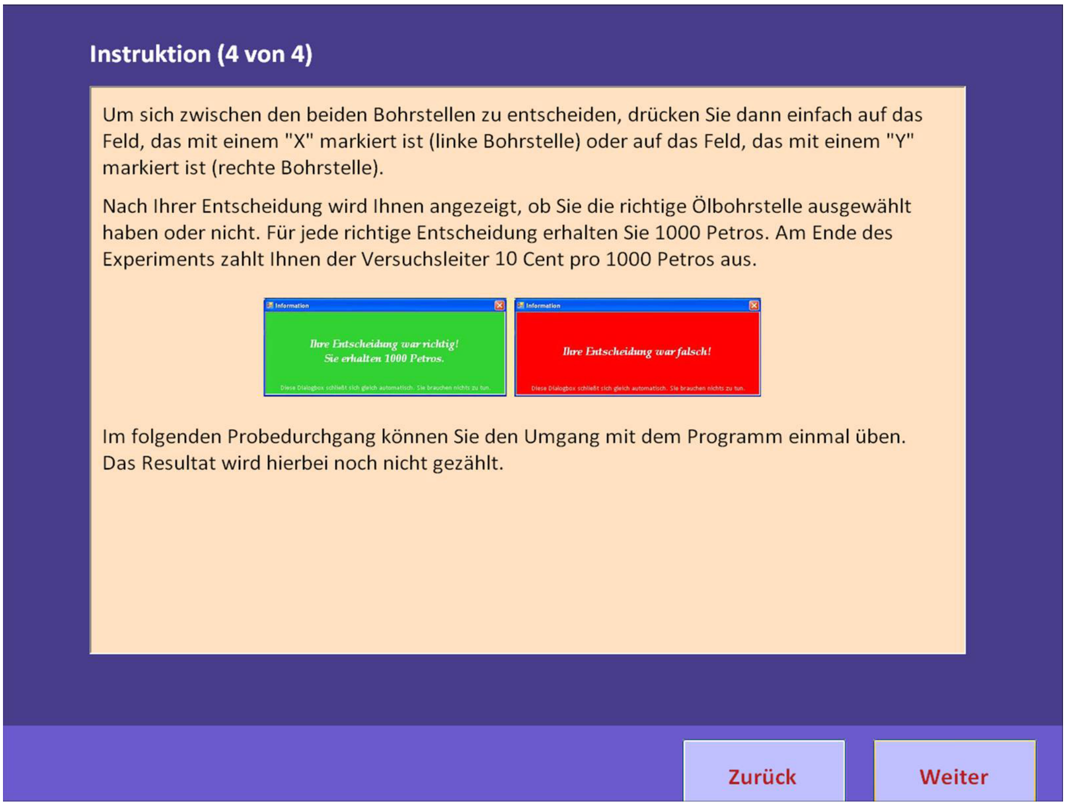
Experimental Material for Chapter 2

Instructions experiment 1 (Note that the original instructions were in German, the translation is given below the screenshots.)

1) Imagine to be a geologist and to have the order of an oil drilling company to find profitable oil drilling sites. In the following, you are supposed to choose the more profitable of two oil-drilling sites. In order to make a decision you can commission six different measures (that is, you can click on them). The six measures can inform you with different levels of certainty ("success") whether one oil drilling site is profitable ("+") or not ("-").

2) See for example the figure below "seismic analyses": If it discriminates, it allows you in 78% of cases a correct prediction about whether you can find oil ("+") or not ("-"). The measure "chemical analyses" in the example below, however, only allows for 53% correct predictions.

3) You are free to choose which and how many measures and in which order you “commission” them (that is, which ones you uncover), until you choose one of the two oil drilling sites (X or Y). To see the result of a measure, just click on the corresponding box with the question mark.



4) In order to choose one of the two oil drilling sites, just click either on the box with the “X” (left oil drilling site) or on the box with the “Y” (right oil drilling site). After your choice, you will receive feedback about the accuracy of your choice. For each correct choice, you will receive 1000 Petros. At the end of the experiment, the experimenter will pay you €0.10 in exchange for 1000 Petros.

In the following practice trial you can practice how the program works. The result is not yet counted.

Additional oral instructions by the experimenter:

5) “Please read through the instructions. There will be a practice trial. If you have questions, please ask me. There is no time limit. [*in dyad condition*: Please work jointly on the task and do not leave it to one person to click on the boxes.]”

Table A.1

Item set in the WADD-friendly environment.

#	Alternative X						Alternative Y						correct
	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6	
1	1	0	0	1	1	1	1	1	1	0	0	0	Y
2	0	1	0	0	0	0	0	0	0	1	1	0	Y
3	1	1	1	0	1	1	1	1	1	1	0	0	X
4	0	0	0	1	1	1	0	0	1	0	1	0	X
5	1	0	0	0	1	1	0	1	1	1	0	0	Y
6	1	0	0	1	1	1	0	1	1	1	1	1	Y
7	0	0	0	0	1	1	1	0	0	0	0	1	Y
8	0	0	0	1	1	1	1	1	1	1	0	0	Y
9	1	0	1	0	0	1	0	1	0	0	1	1	X
10	0	0	0	1	1	1	1	1	1	0	0	0	Y
11	1	1	0	1	0	1	0	0	1	1	0	0	X
12	0	0	1	0	1	0	1	1	1	1	0	0	Y
13	0	0	1	1	0	0	1	0	0	0	1	0	Y
14	0	1	1	1	1	1	1	1	0	1	0	1	Y
15	1	0	1	1	1	0	0	0	0	0	0	1	X
16	1	0	0	0	0	1	0	1	1	0	0	0	Y
17	1	1	0	0	1	1	1	1	1	0	0	0	X
18	0	0	1	0	1	0	1	0	0	0	1	1	Y
19	0	1	0	0	0	0	0	0	1	0	1	0	Y
20	0	0	0	0	0	1	1	1	1	1	0	0	Y
21	0	0	0	1	1	1	0	1	0	0	0	0	X
22	0	0	0	1	1	1	1	1	1	0	1	1	Y
23	0	0	0	1	0	1	1	1	1	0	1	1	Y
24	1	0	0	0	0	0	0	0	0	1	0	1	Y
25	1	0	0	0	1	0	0	1	1	0	0	0	X
26	1	1	1	1	1	1	1	0	0	0	0	0	X

Note. C1 = cue 1, C2 = cue 2, etc.; correct = correct alternative.

Table A.2

Item set in the take-the-best-friendly environment.

#	Alternative X						Alternative Y						correct
	C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6	
1	1	0	0	1	1	1	1	1	1	0	0	0	Y
2	0	0	0	0	1	0	0	1	1	1	0	0	Y
3	0	0	1	0	1	0	0	0	0	1	0	1	X
4	0	1	0	0	1	1	0	0	0	0	0	0	X
5	1	1	1	1	0	0	1	1	1	1	1	1	Y
6	0	0	0	1	1	1	0	0	1	0	1	0	X
7	1	0	0	0	0	1	0	0	0	1	1	1	X
8	0	1	0	1	1	1	0	1	1	1	0	0	Y
9	1	1	1	1	1	1	1	0	0	0	0	0	X
10	1	0	0	0	0	1	1	1	1	1	1	1	Y
11	0	1	0	1	1	1	0	0	1	1	0	0	X
12	1	0	1	1	1	0	0	1	1	1	1	1	X
13	1	0	0	0	1	0	0	1	0	1	1	1	X
14	1	0	0	0	0	0	0	0	0	1	0	1	Y
15	1	0	0	0	0	0	1	1	0	1	0	1	Y
16	0	0	0	0	0	0	0	0	0	1	1	1	Y
17	1	0	0	1	1	1	0	1	1	1	1	1	Y
18	1	0	1	1	1	1	1	1	0	0	1	1	Y
19	0	0	0	0	1	1	0	0	1	1	0	0	Y
20	1	0	1	1	1	0	0	1	0	0	0	0	X
21	1	0	0	0	0	1	0	0	0	0	1	1	X
22	0	0	1	0	1	0	0	0	0	1	1	1	X
23	1	1	0	1	0	1	1	0	1	1	1	1	X
24	1	1	0	1	0	1	0	1	1	1	1	1	X
25	0	0	1	0	1	0	0	0	0	0	0	0	X
26	0	1	1	0	0	0	1	0	0	0	1	0	Y

Note. C1 = cue 1, C2 = cue 2, etc.; correct = correct alternative.

Appendix B

Additional Results for Chapter 2

Accordance rates with take-the-best and WADD in Experiment 1

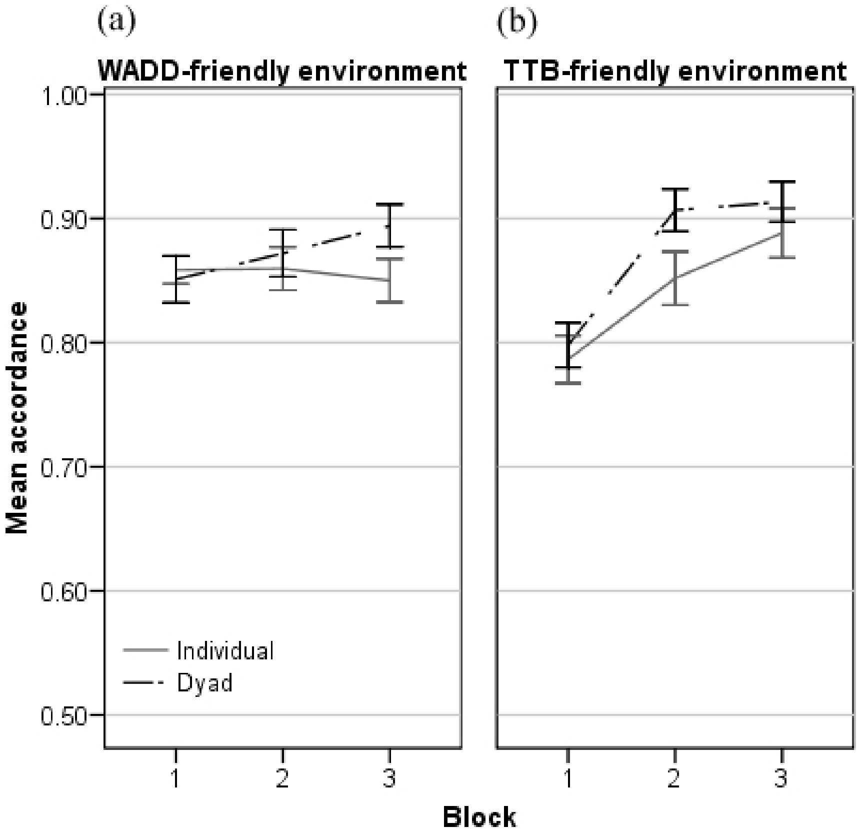


Figure B.1. Individuals' and dyads' mean rates of accordance with the adaptive strategy in the (a) WADD- friendly and (b) take-the-best- (TTB-) environments. In both environments, choices were strongly in accordance with the appropriate adaptive strategy. Dyads, however, either reached asymptotic accordance faster (take-the-best-friendly environment) or reached higher final levels of accordance with the adaptive strategy (WADD-friendly environment). Error bars: ±1 *SE*.

Accordance rates with take-the-best and WADD in Experiment 2

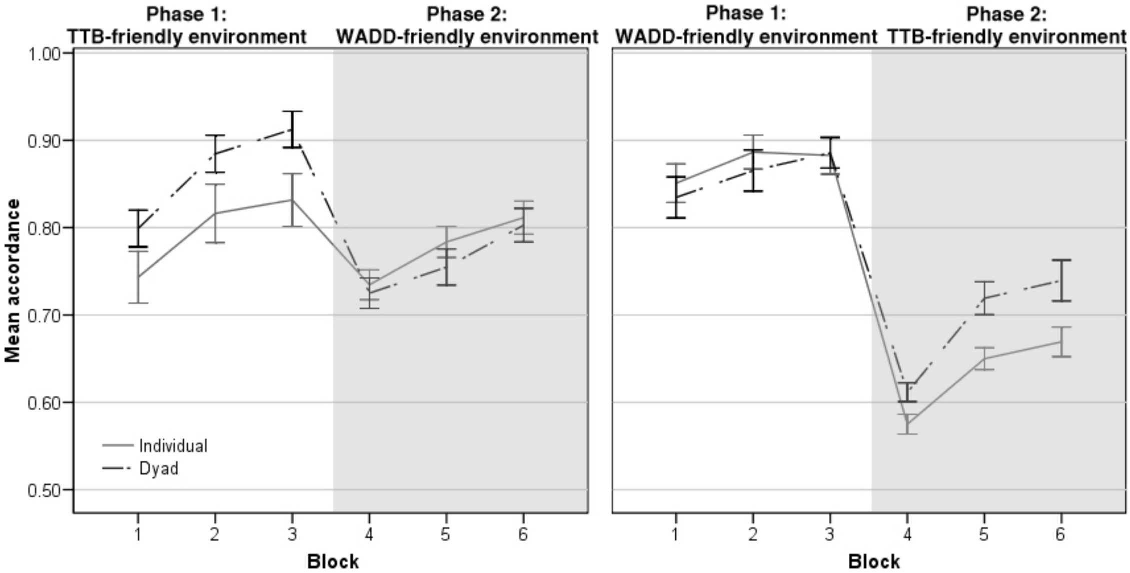


Figure B.2. Individuals' and dyads' mean accordance rates with the adaptive strategy in the WADD-friendly and the take-the-best-friendly environment, respectively. The two left panels depict the rates of accordance with the adaptive strategies in the experimental order of first the take-the-best-friendly and then the WADD-friendly environment; n = 20 individuals, n = 20 dyads); the two right panel depict the results for the reverse order. Error bars: ±1 *SE*.

Appendix C

Experimental Material Used in Chapter 3

Companies used in main study

Table C.1

100 companies with market capitalization and numbe of times they were recognized by participants (N = 143).

#	Name of company	Stock	Monthly market capitalization of May 2008	Times recognized
1	adidas AG	DAX	9.086.050.220	127
2	Allianz SE	DAX	55.398.666.979	124
3	BASF SE	DAX	43.785.434.550	106
4	Bayer AG	DAX	42.367.936.968	128
5	BMW AG St	DAX	11.662.149.096	121
6	Commerzbank AG	DAX	13.351.642.721	127
7	Continental AG	DAX	12.757.207.715	113
8	Daimler AG	DAX	45.295.009.517	126
9	Deutsche Bank AG	DAX	37.428.036.733	128
10	Deutsche Börse AG	DAX	18.994.391.052	95
11	Deutsche Lufthansa AG	DAX	7.580.969.950	127
12	Deutsche Post AG	DAX	17.561.071.533	127
13	Deutsche Postbank AG	DAX	4.862.009.834	128
14	Deutsche Telekom AG	DAX	30.394.620.101	127
15	E.ON AG	DAX	82.786.198.005	124
16	Fresenius Medical Care AG u. Co. KGaA St	DAX	6.466.279.155	59
17	Henkel KGaA Vz	DAX	5.316.563.030	95
18	Hypo Real Estate Holding AG	DAX	4.337.522.867	43
19	Infineon Technologies AG	DAX	4.843.179.352	89
20	Linde AG	DAX	13.317.926.271	51
21	MAN AG St	DAX	9.677.891.582	73
22	Merck KGaA	DAX	5.798.774.688	38
23	METRO AG St	DAX	5.435.947.348	105
24	Münchener Rück AG	DAX	24.995.380.725	61
25	RWE AG St	DAX	32.587.464.970	91
26	SAP AG	DAX	29.356.960.833	79
27	Siemens AG	DAX	64.300.621.507	129
28	ThyssenKrupp AG	DAX	15.193.530.522	124

Continuation of Table C.1

#	Name of company	Stock	Monthly market capitalization of May 2008	Times recognized
29	TUI AG	DAX	2.771.285.676	123
30	Volkswagen AG St	DAX	26.668.951.484	123
31	Aareal Bank AG	MDAX	586.495.578	11
32	Altana AG	MDAX	876.970.929	25
33	AMB Generali Holding AG	MDAX	895.928.544	26
34	Beiersdorf Aktiengesellschaft	MDAX	4.176.217.899	66
35	Bilfinger Berger AG	MDAX	1.967.648.626	17
36	Celesio AG	MDAX	2.147.277.649	7
37	Deutsche EuroShop AG	MDAX	750.498.436	15
38	Deutz AG	MDAX	417.071.720	28
39	Douglas Holding AG	MDAX	1.011.489.327	90
40	EADS N.V.	MDAX	5.043.414.881	36
41	Fraport AG	MDAX	1.564.413.614	33
42	Fresenius SE Vz	MDAX	4.221.523.942	62
43	GAGFAH S.A.	MDAX	777.963.718	33
44	Gildemeister AG	MDAX	913.593.061	10
45	Hamburger Hafen und Logistik AG	MDAX	1.209.407.764	46
46	Hannover Rückversicherung AG	MDAX	2.068.606.711	65
47	HeidelbergCement AG	MDAX	1.923.376.964	21
48	Heidelberger Druckmaschinen AG	MDAX	1.013.408.244	26
49	HOCHTIEF AG	MDAX	3.265.147.639	80
50	Hugo Boss AG Vz	MDAX	499.346.368	120
51	IVG Immobilien AG	MDAX	1.462.554.580	8
52	K+S Aktiengesellschaft	MDAX	10.117.589.360	21
53	Krones AG	MDAX	830.686.782	10
54	KUKA Aktiengesellschaft	MDAX	576.625.506	2
55	LANXESS AG	MDAX	2.313.128.502	4
56	Leoni AG	MDAX	992.362.466	4
57	MLP AG	MDAX	711.533.834	23
58	MTU Aero Engines Holding AG	MDAX	1.507.075.922	17
59	Norddeutsche Affinerie AG	MDAX	1.009.863.775	9
60	Praktiker Bau- und Heimwerkermärkte Holding AG	MDAX	864.447.827	98
61	Premiere AG	MDAX	1.040.496.799	117
62	ProSiebenSat.1 Media AG	MDAX	813.351.455	127
63	Puma AG	MDAX	1.419.332.559	121
64	Rheinmetall AG	MDAX	1.857.805.740	63
65	Salzgitter AG	MDAX	5.406.853.547	43
66	STADA Arzneimittel AG	MDAX	2.546.932.786	64
67	Symrise AG	MDAX	1.683.956.221	4
68	Vossloh AG	MDAX	935.096.637	15
69	Wacker Chemie AG	MDAX	2.792.828.952	29
70	WINCOR NIXDORF Aktiengesellschaft	MDAX	1.663.843.005	53

Continuation of Table C.1

#	Name of company	Stock	Monthly market capitalization of May 2008	Times recognized
71	Air Berlin PLC	SDAX	332.842.549	115
72	Axel Springer AG	SDAX	603.620.768	128
73	BAUER Aktiengesellschaft	SDAX	526.959.928	67
74	BayWa AG vNa	SDAX	625.724.016	13
75	Biotest AG Vz	SDAX	220.000.016	6
76	C.A.T. OIL AG	SDAX	186.826.349	12
77	Colonia Real Estate AG	SDAX	198.896.196	14
78	comdirect bank AG	SDAX	247.264.649	107
79	CTS Eventim AG	SDAX	272.585.049	47
80	Deutsche Beteiligungs AG	SDAX	232.134.135	12
81	Deutsche Wohnen AG	SDAX	332.778.566	15
82	Dyckerhoff AG Vz	SDAX	185.280.767	17
83	Duerr AG	SDAX	185.098.676	18
84	elexis AG	SDAX	173.070.159	8
85	EM.Sport Media AG	SDAX	138.946.637	17
86	Escada AG St	SDAX	152.304.989	59
87	Fielmann AG	SDAX	591.221.110	126
88	Gerresheimer AG	SDAX	1.075.208.389	5
89	Gerry Weber International AG	SDAX	240.185.826	85
90	GfK AG	SDAX	442.378.885	17
91	Homag Group AG	SDAX	216.535.583	9
92	IKB Dt. Industriebank AG	SDAX	167.861.461	32
93	Jungheinrich AG	SDAX	360.550.645	17
94	KOENIG u. BAUER AG	SDAX	275.692.738	13
95	Medion AG	SDAX	301.768.021	95
96	MPC AG	SDAX	279.948.423	3
97	MVV Energie AG	SDAX	386.052.273	7
98	PATRIZIA Immobilien AG	SDAX	100.941.435	2
99	Sixt AG St	SDAX	238.013.743	97
100	Vivacon AG	SDAX	169.378.348	24

Instructions of Main Study

Introduction

Herzlich willkommen zu unserer Studie zum Thema "Bekanntheit von börsennotierten Unternehmen in Deutschland". Im Folgenden werden Ihnen drei Teilaufgaben präsentiert, mit deren Hilfe wir erfassen möchten, welche und inwiefern Ihnen an der Deutschen Börse notierte Unternehmen bekannt sind. Dazu möchten wir Sie bitten, die Fragen ehrlich, vollständig und intuitiv zu beantworten. Alle Angaben bleiben anonym und werden ausschließlich zu internen Forschungszwecken verwendet.

Wir bedanken uns für Ihre Teilnahme!

Drücken Sie die Leertaste um mit der Studie zu beginnen

Recognition Task

Im Folgenden wird Ihnen eine Liste mit den Namen von 100 Unternehmen (in zufälliger Reihenfolge) dargeboten, die an der Deutschen Börse (DAX, MDAX, SDAX) notiert sind. Uns interessiert, von welchen dieser Unternehmen Sie schon einmal gehört haben. Zunächst ist uns nicht wichtig, ob Sie auch Näheres über das Unternehmen wissen, nur, ob Sie schon einmal davon gehört haben oder nicht.

Bitte antworten Sie ohne langes Überlegen so schnell wie möglich, ob Sie von dem gezeigten Unternehmen schon einmal etwas gehört haben oder nicht.

Bitte verwenden Sie für die Beantwortung der Fragen den linken und den rechten Zeigefinger. Legen Sie vor der Bearbeitung jeder Frage diese Finger auf die entsprechenden Antworttasten:

Q (NEIN, habe noch nie von dem Unternehmen gehört, linker Zeigefinger)
P (JA, habe schon einmal von dem Unternehmen gehört, rechter Zeigefinger)

Bei jeder Frage wird an Stelle des jeweiligen Unternehmens zunächst nur kurz ein Kreuz sichtbar sein. Sobald das Kreuz verschwindet und das jeweilige Unternehmen erscheint, können Sie durch einen Tastendruck antworten.\n\n\nBeachten Sie dabei bitte, dass Sie Ihre Antwort nicht mehr ändern können, nachdem Sie sich bei einer Frage durch einen Tastendruck einmal entschieden haben.

Drücken Sie die Leertaste für weitere Instruktionen.

Für den Fall, dass Sie von dem Unternehmen schon einmal gehört haben, fragen wir Sie einem nächsten Schritt, ob Sie über dieses Unternehmen auch noch mehr Wissen haben, oder ob Sie nur davon gehört haben, ohne mehr darüber zu wissen.

Bitte verwenden Sie auch hier wieder die Tasten Q und P:
Q (NEIN, kenne nur den Namen, linker Zeigefinger)
P (JA, weiß noch mehr darüber, rechter Zeigefinger)
Drücken Sie die Leertaste um mit der Aufgabe zu beginnen.

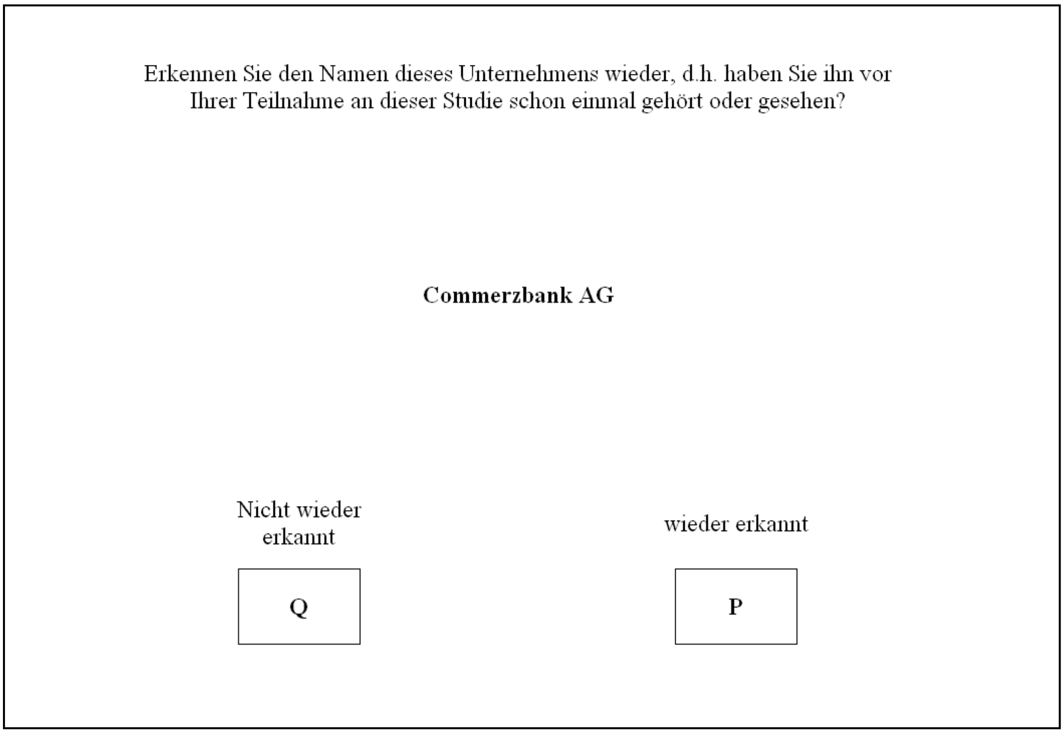


Figure C.1. Example of a recognition-test slide (recognized or not recognized).

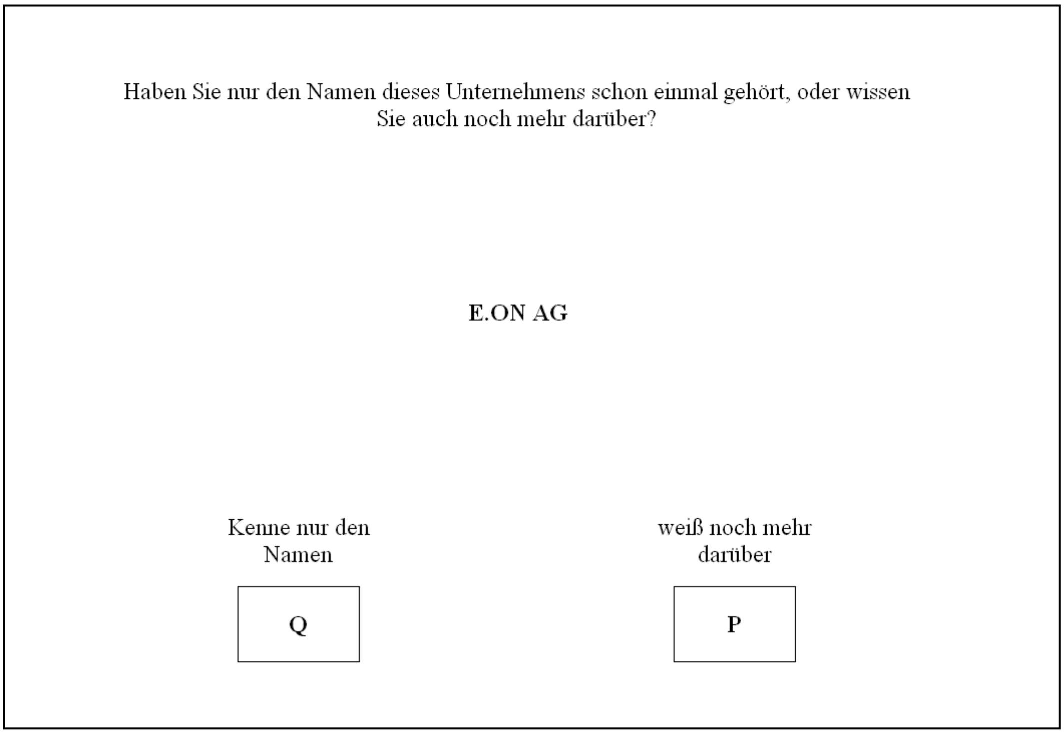


Figure C.2. Example of a recognition-test slide (if recognized whether the person has more information or not).

Individual Paired Comparison Task

Im nun folgenden Teil unserer Studie werden Ihnen immer zwei Unternehmensnamen gleichzeitig auf dem Bildschirm dargeboten. Ihre Aufgabe ist es, so schnell wie möglich zu entscheiden, welches der beiden Unternehmen zurzeit einen höheren Börsenwert hat. \n\nZu Ihrer Information: Der Börsenwert, auch Markt-/Börsenkapitalisierung genannt, spiegelt den aktuellen Gesamtwert eines Unternehmens wider und entspricht dem Preis, den ein Käufer für sämtliche umlaufenden Aktien eines Unternehmens - also eine komplette Übernahme – bezahlen müsste. Er wird z.B. berechnet, indem man die Anzahl der Aktien mit dem Aktienkurs multipliziert. Bitte schätzen Sie intuitiv ein, welches der gezeigten Unternehmen momentan mehr wert ist.

Es werden Ihnen 50 Paarvergleiche dargeboten, jedoch werden Sie nach den Paarvergleichen kein Feedback bekommen.

Drücken Sie die Leertaste für weitere Instruktionen.

Bitte verwenden Sie für die Beantwortung der Fragen den linken und den rechten Zeigefinger. Legen Sie vor der Bearbeitung jeder Frage diese Finger auf die entsprechenden Antworttasten:

Q (Unternehmen auf der linken Seite, linker Zeigefinger)
P (Unternehmen auf der rechten Seite, rechter Zeigefinger)

Bei jeder Frage werden an Stelle der jeweiligen Unternehmen zunächst kurz Kreuze sichtbar sein. Sobald die Kreuze verschwinden und die jeweiligen Unternehmen erscheinen, können Sie durch einen Tastendruck antworten.

Beachten Sie dabei bitte, dass Sie Ihre Antwort nicht mehr ändern können, nachdem Sie sich bei einer Frage durch einen Tastendruck einmal entschieden haben.

Nach jeder Entscheidung werden Sie gefragt, wie sicher Sie sich jeweils sind. Geben Sie hierzu dann bitte Zahlen zwischen 1 (sehr unsicher) und 5 (sehr sicher) ein. Mit Zahlen dazwischen können Sie Ihr Urteil abstufen.

Drücken Sie die Leertaste um mit der Aufgabe zu beginnen.

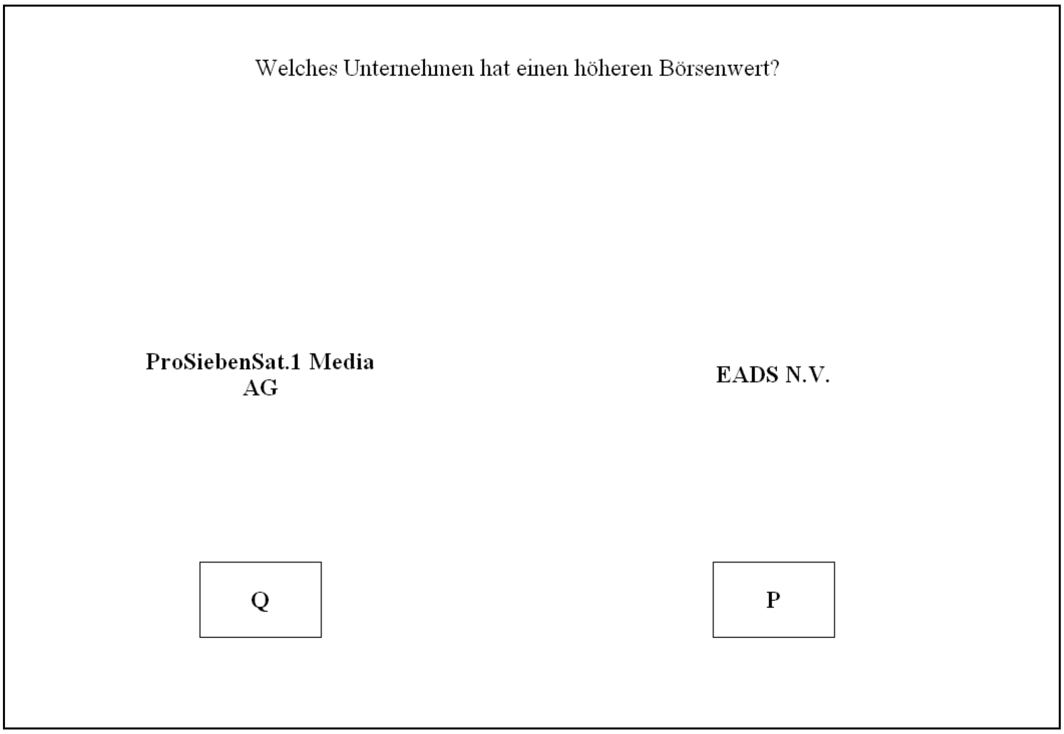


Figure C.3. Example of an individual paired comparison-test slide (choice between two companies).

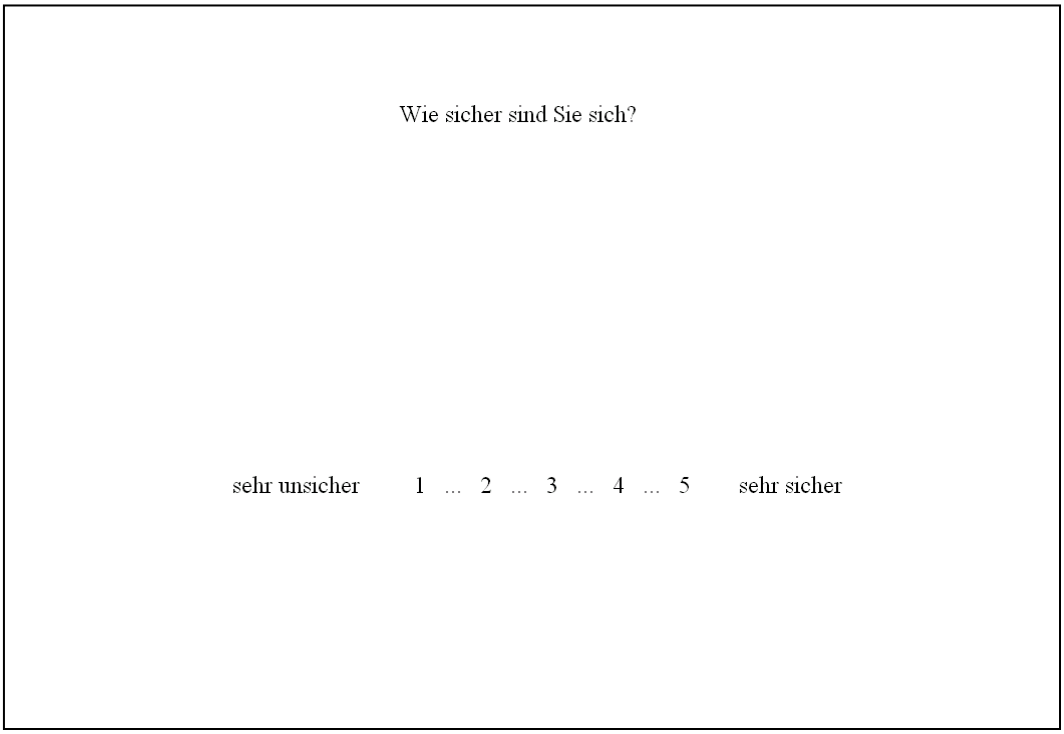


Figure C.4. Example of a paired comparison-test slide (indication of certainty level after the choice).

Group paired comparison task

Without time pressure

Herzlich willkommen zum Gruppenteil der Aufgabe. Bitte entscheiden Sie nun gemeinsam, welches der beiden Unternehmen den höheren Börsenwert hat. Zur Erinnerung:

Q (Unternehmen auf der linken Seite)
P (Unternehmen auf der rechten Seite)

Bei jeder Frage werden an Stelle der jeweiligen Unternehmen zunächst kurz Kreuze sichtbar sein. Sobald die Kreuze verschwinden und die jeweiligen Unternehmen erscheinen, können Sie durch einen Tastendruck antworten. Beachten Sie dabei bitte, dass Sie Ihre Antwort nicht mehr ändern können, nachdem Sie sich bei einer Frage durch einen Tastendruck einmal entschieden haben.

Drücken Sie die Leertaste um mit der Aufgabe zu beginnen.

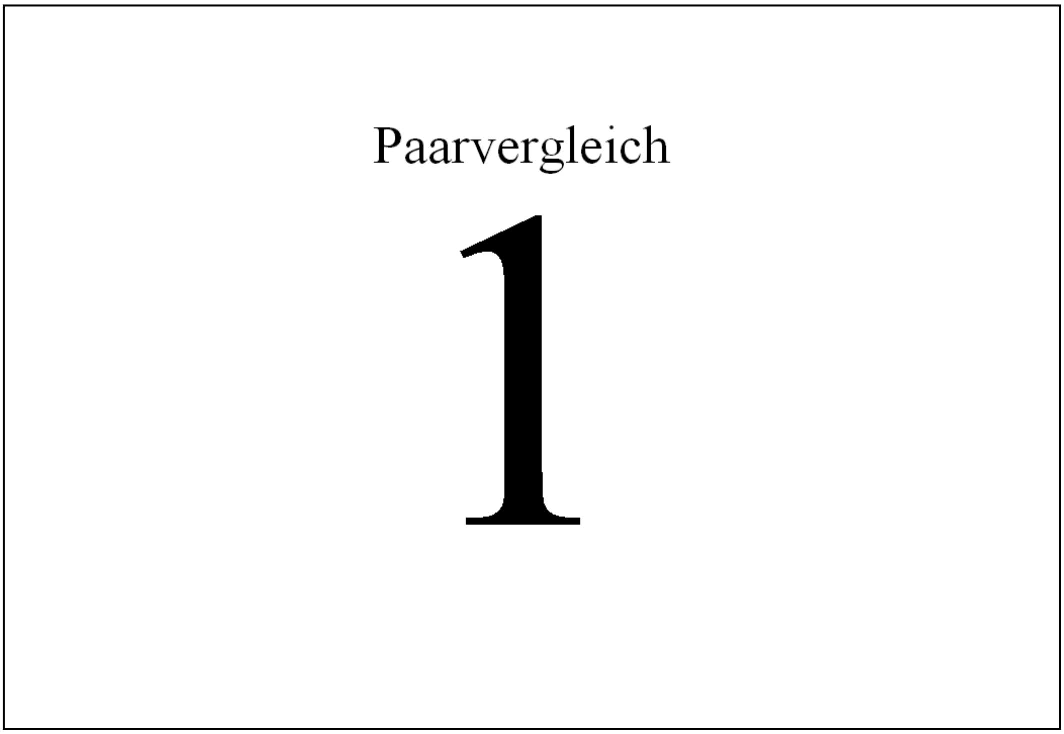


Figure C.5. Example of a group paired comparison-test slide (indicating a new comparison).

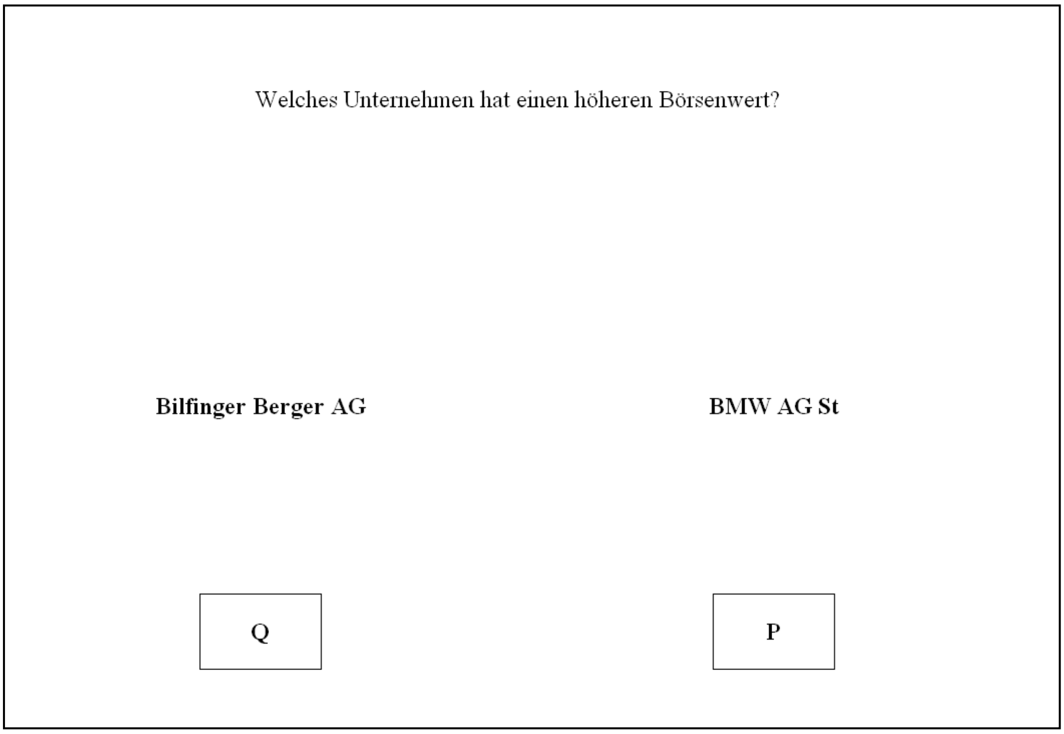


Figure C.6. Example of a group paired comparison-test slide (choice between two companies, without countdown, condition without time pressure).

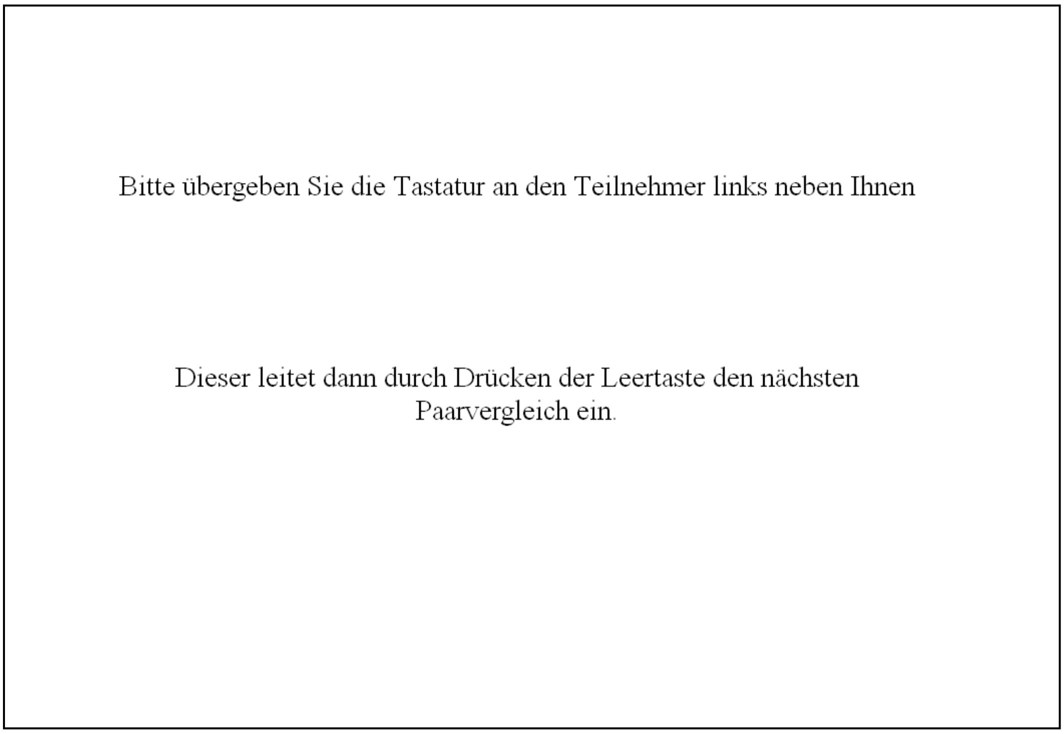


Figure C.7. Example of a group paired comparison-test slide (slide between two comparisons).

With time pressure

Herzlich willkommen zum Gruppenteil der Aufgabe. Bitte entscheiden Sie nun gemeinsam, welches der beiden Unternehmen den höheren Börsenwert hat. Zur Erinnerung:

Q (Unternehmen auf der linken Seite)
P (Unternehmen auf der rechten Seite)

Bei jeder Frage werden an Stelle der jeweiligen Unternehmen zunächst kurz Kreuze sichtbar sein. Sobald die Kreuze verschwinden und die jeweiligen Unternehmen erscheinen, können Sie durch einen Tastendruck antworten. Beachten Sie dabei bitte, dass Sie Ihre Antwort nicht mehr ändern können, nachdem Sie sich bei einer Frage durch einen Tastendruck einmal entschieden haben.

Sie haben jeweils 30 Sekunden Zeit, was Ihnen auch durch einen Countdown angezeigt wird.

Drücken Sie die Leertaste um mit der Aufgabe zu beginnen.

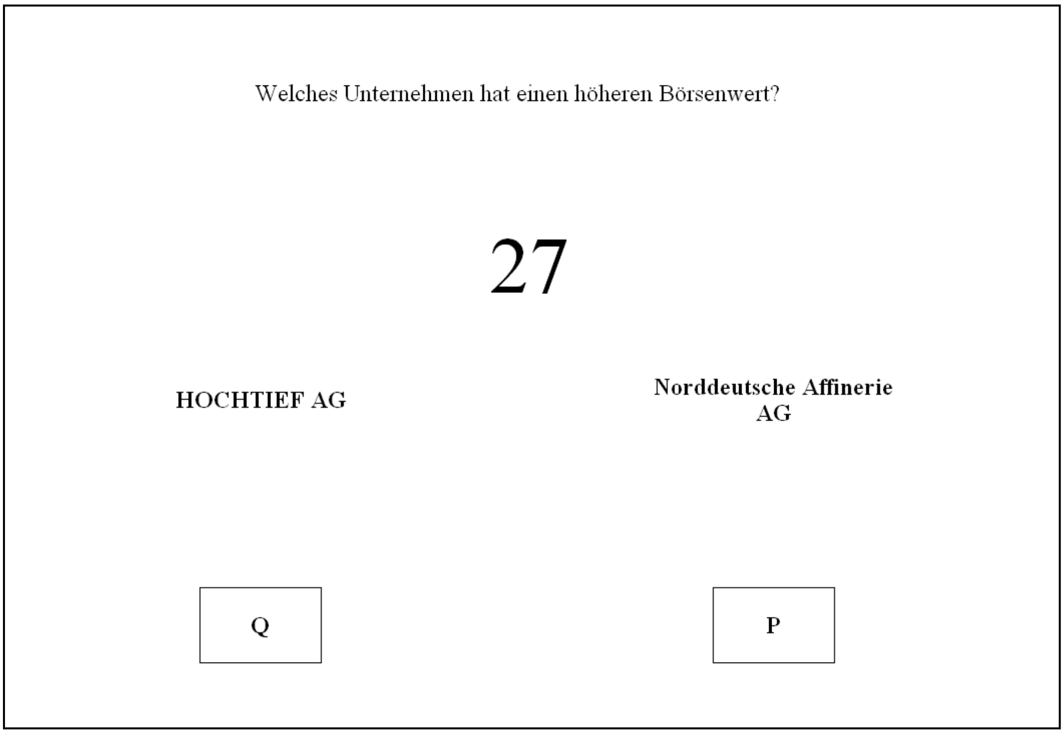


Figure C.8. Example of a group paired comparison-test slide (choice between two companies, with countdown, condition with time pressure).

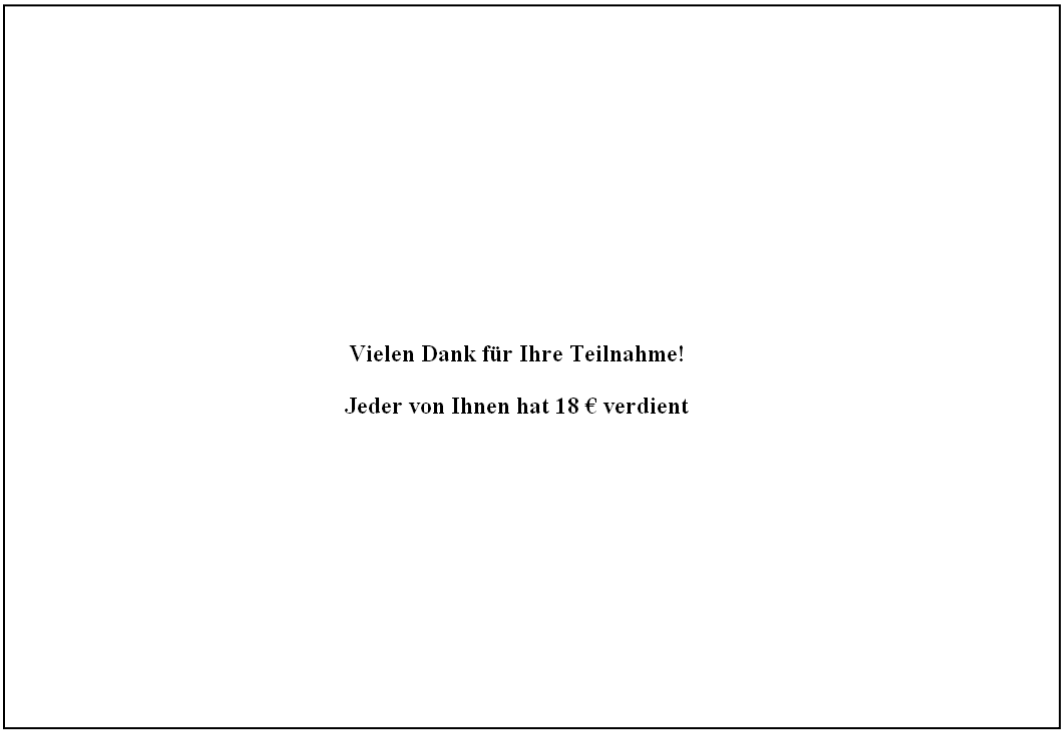


Figure C.9. Last slide of group task indicating the reward for each participant.

Argument Recall Task

Bekanntheit von börsennotierten Unternehmen in Deutschland
Kämmer, J., Gaissmaier, W. & Schermuly, C.

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Fragebogen 1 2 3 4 „Einschätzung der Diskussion"

An welche wichtigen Argumente erinnern Sie sich, die für und gegen einen hohen Börsenwert eines Unternehmens sprechen?
*Bitte tragen Sie die für Sie überzeugendsten Argumente in die Tabelle ein. Ordnen Sie Argumente, die für einen hohen Börsenwert sprachen, entsprechend ihrer Wichtigkeit in die linke Spalte ein (**1 = am wichtigsten**). Tragen Sie Argumente, die gegen einen hohen Börsenwert sprachen, in die rechte Spalte ebenfalls entsprechend ihrer Wichtigkeit geordnet ein.*

Das spricht **FÜR** einen hohen Börsenwert:	Das spricht **GEGEN** einen hohen Börsenwert:
1. ______	1. ______
2. ______	2. ______
3. ______	3. ______
4. ______	4. ______

Datum ______ Teilnehmer A

Figure C.10. Page of argument recall task (recall of the most important arguments speaking for and against a high market capitalization).

Rules for Coding Discussion Behavior with DCS (Schermuly & Scholl, 2012)

What to code

All discussion behaviours between the releasing a new trial and finishing a trial (by entering the decision) were coded, but not the interaction between two trials.

Categories

Recognition cue

An act was coded as recognition cue when a person used the recognition heuristic as an argument or simply let the other members know that she/he recognized one company name but not the other. It was also coded when a person used the recognition cue of another member or the whole group as an argument. Thus, it is coded when (un-)recognition is given as a reason for a decision.

Examples:

"If you do not even recognize company A, it cannot be big."

"Let's take company A, since we all recognize it."

"I only recognize company A."

"I chose A because I don't know B."

Knowledge cue

Acts were coded as knowledge cues when a group member provided cues about a company other than recognition. This comprises situations in which a person simply stated his/her cue knowledge or used a cue as an argument for or against a high market capitalization and thus as a reason for a decision.

Examples:

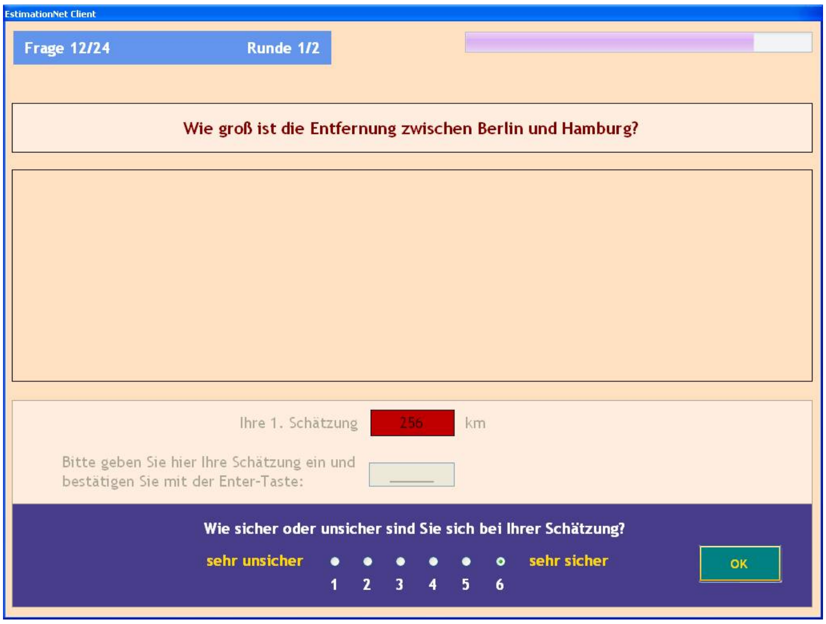
"I know that company B produces drugs."

"Company A is a bank, and banks have money."

Appendix D

Experimental Material Used in Chapter 4

(a)



(b)

Figure D.1. Screenshots of the advice condition. Participants were first asked for their initial estimate (a). Before giving their second estimate, they received the estimate of another randomly drawn person from the prestudies and this person's confidence (b).

Table D.1

Items, true answers and mean confidence ratings of the prestudies and the main study.

#	Item	Truth	Confidence Mean (SD) Main study	Confidence Mean (SD) Prestudies
	easy			
1	What is the maximum speed of a cheetah (in kilometres per hour)?	110	3.61 (1.40)	3.58 (1.25)
2	What is the average height of a 10-years old boy (in Germany, in centimetres)?	141	3.40 (0.98)	3.32 (0.99)
3	What is the monthly amount of child benefit in Germany for the first child (in 2012)?	184	4.19 (1.34)	3.75 (1.45)
4	What is the distance between Berlin and Hamburg (in kilometres)?	256	3.79 (1.19)	4.31 (1.12)
5	How many feet (English measure of length) equal 100 meters?	328	2.63 (1.34)	2.88 (1.39)
6	What is the monthly amount of basic security benefits for full-aged, single job seekers (Hartz IV, in 2012)?	364	3.66 (1.38)	3.93 (1.35)
7	What is the height of the Fernsehturm in Berlin (in meters)?	368	3.28 (1.41)	3.67 (1.50)
8	How much does the iPad3 with 16GB cost (RRP)?	479	3.09 (1.31)	3.14 (1.38)
	intermediate			
9	How many movie theatres are there in Berlin (retrieved 2008)?	170	2.69 (1.21)	2.66 (1.12)
10	How many countries take part in the general assembly of the United Nations as active members?	193	2.64 (1.38)	2.73 (1.37)
11	How many active nuclear power stations are there in Europe (retrieved 2011)?	196	2.49 (1.18)	2.49 (1.09)
12	What is the world record in high jump of men (in centimetres)?	245	2.77 (1.24)	2.32 (1.21)
13	What is the world record in ski jumping of men (in m)?	247	2.37 (1.28)	2.37 (1.28)
14	What is the length of the border between Switzerland and Germany (in kilometres)?	316	2.66 (1.26)	2.53 (1.29)
15	What is the speed of sound in the air (on sea level, in meter per second)?	343	2.27 (1.62)	2.55 (1.69)
16	What is the length of the river Oder in kilometres?	866	2.17 (1.07)	2.21 (1.02)
	difficult			
17	How many earthquakes with a value of more than 6 on the Richter scale happen in an average year worldwide?	150	2.22 (1.06)	2.09 (1.08)
18	How many athletes took part in the first modern Olympic Games in Athens in 1896?	241	1.81 (0.96)	1.60 (0.80)

Table D.1 (continued)

#	Item	Truth	Confidence Mean (SD)	
			Main study	Prestudies
19	Of yow many days does a year according to the ritual Maya-calendar comprise?	260	1.89 (1.15)	1.37 (0.55)
20	How deep is the Baltic Sea at its deepest point (in meters)?	459	2.08 (1.04)	1.94 (0.99)
21	How many people have the Vatican citizenship (retrieved 2011)?	572	2.00 (1.12)	1.80 (1.00)
22	How many (earth)days has a year on the Mars?	687	1.84 (0.97)	1.75 (1.18)
23	How many accomplished murders and homicides were officially registered in Germany in 2010?	690	2.18 (1.04)	1.81 (0.97)
24	How many inhabitants has the East Frisian island Wangerooge (retrieved 2010)?	919	1.92 (0.95)	1.86 (1.01)

Note. The average confidence ratings per difficulty found in the prestudies were (with SD in parentheses): for easy items: 3.57 (1.13), for intermediate items: 2.48 (1.00), for difficult items: 1.78 (0.76).

Appendix E

Additional Results for Chapter 4

(a)

(b)

(c)

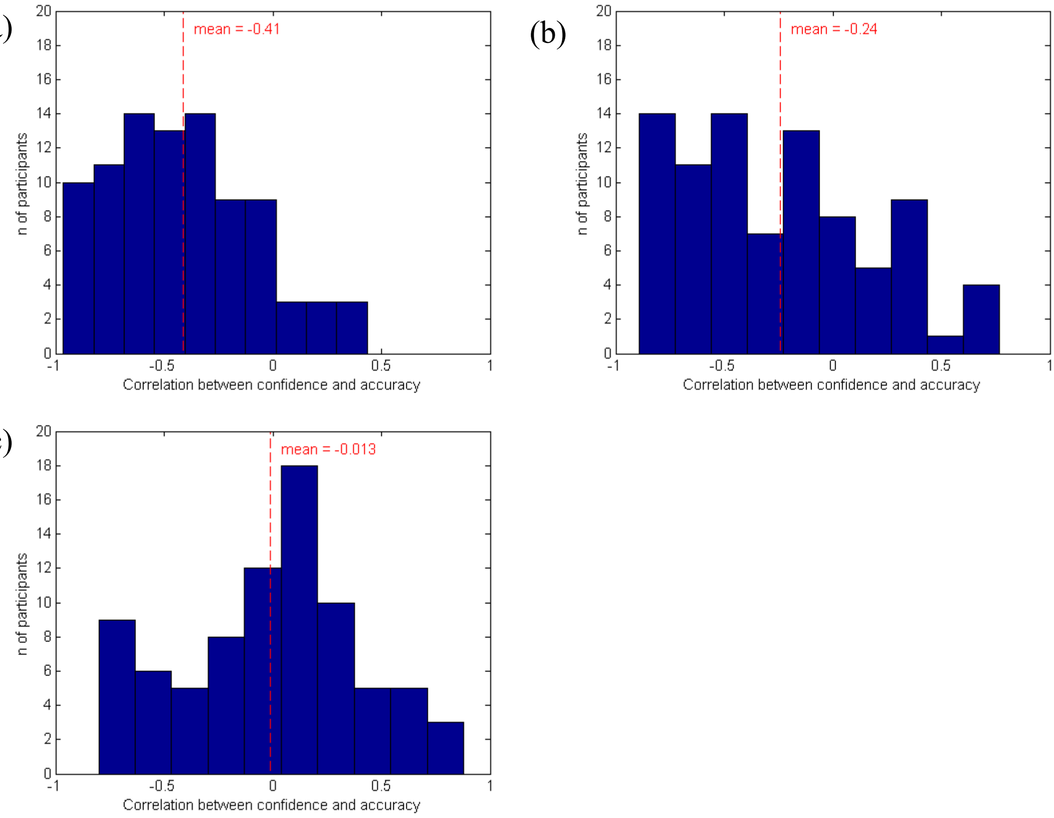


Figure E.1. Histogram of intraindividual correlations (Spearman's *rho*) between confidence and error (e_1) for (a) easy, (b) intermediate, and (c) difficult tasks.

Table E.1

Results of paired t-tests for comparisons within environments.

Environment	Comparison	Paired t-test result
easy	$ws = 0$ vs. $ws = 1$	$t(29) = -0.438$, $p = .664$, $dz = 0.08$
	$ws = 0$ vs. $ws = .5$	$t(29) = -0.930$, $p = .360$, $dz = 0.17$
	$ws = 1$ vs. $ws = .5$	$t(29) = 0.574$, $p = .570$, $dz = 0.11$
intermediate	$ws = 0$ vs. $ws = 1$	$t(29) = -0.396$, $p = .695$, $dz = 0.07$
	$ws = 0$ vs. $ws = .5$	$t(29) = 2.945$, $p = .006$, $dz = 0.54$
	$ws = 1$ vs. $ws = .5$	$t(29) = 2.975$, $p = .006$, $dz = 0.54$
difficult	$ws = 0$ vs. $ws = 1$	$t(29) = -1.500$, $p = .144$, $dz = 0.27$
	$ws = 0$ vs. $ws = .5$	$t(29) = 4.517$, $p < .001$, $dz = -0.24$
	$ws = 1$ vs. $ws = .5$	$t(29) = 2.679$, $p = .012$, $dz = 0.49$

Note. $ws = 0$ = choose the other, $ws = .5$ = average, $ws = 1$ = choose the self

List of Tables

List of Figures

Chapter 2

Chapter 3

Chapter 4

Appendix B

Appendix C

Appendix D

Appendix E

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig und ohne die unzulässige Hilfe Dritter verfasst habe und die Dissertation auch in Teilen keine Kopie anderer Arbeiten darstellt. Die verwendeten Hilfsmittel sowie Literatur sind vollständig angegeben. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden. Ich habe keinen Doktorgrad in dem Promotionsfach Psychologie, und die zugrundeliegende Promotionsordnung der Humboldt-Universität vom 03.08.2006 ist mir bekannt.

Berlin, den 14. Dezember 2012

____
Juliane Eva Kämmer